The London School of Economics and Political Science

# Essays on Moral Doubt, Meta-ethics, and Choice

## Nicholas Makins

A thesis submitted to the Department of Philosophy, Logic and Scientific Method of the London School of Economic and Political Science for the degree of Doctor of Philosophy, London, January 2022.

# Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work.

The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without my prior written consent. I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I confirm that an edited version of Chapter 3 is forthcoming as "Attitudinal Ambivalence: Moral Uncertainty for Non-cognitivists" in the *Australasian Journal of Philosophy*.

I declare that my thesis consists of 50,336 words.

# Abstract

This thesis is comprised of four papers that explore the nature and practical implications of moral doubt. It starts from the observation that the appropriate way to understand and handle moral doubt will depend on the nature of our moral judgements themselves. The first and second papers take a cognitivist view of moral judgements and analyse moral doubt in terms of partial degrees of belief in moral propositions. The first, The Relevance of Belief: Subjective Norms Under Empirical and Moral Uncertainty, argues that this form of moral uncertainty is relevant to our choices and that the norms most directly relevant to choice under moral uncertainty are the action-guiding subjective norms of instrumental rationality. The second, Preference and Commitment Under Moral Uncertainty, proposes a particular way of accommodating moral uncertainty within a decision theoretic model of rational choice, and shows that this has substantial advantages over existing alternatives. The third paper, Attitudinal Ambivalence: Moral Uncertainty for Non-cognitivists, provides an account of moral doubt that is consistent with non-cognitivism, by analysing the phenomenon in terms of ambivalence in one's moral attitudes. The fourth and final paper, The Balance and Weight of Reasons, draws out some implications of ambivalence, by exploring the way in which the overall weight of reasons are reflected in the preferences of rational agents.

# Acknowledgements

# Table of Contents

# Introduction

You are walking to meet a friend for a coffee, when a person stops you, tells you that they are homeless and hungry, and asks for some money to get something to eat. You want to do the right thing. But, at that moment, you are unsure of just what the right thing is. On the one hand, you have some cash in your pocket that could help this person here and now. On the other, directly handing out money will do nothing to alter the social and political structures that give rise to hunger and homelessness in the first place. You are unsure of whether to help the person who stands in front of you, or instead to direct your resources to support the work of charities, such as Crisis and Shelter, in their campaigns for better housing and social policy.

Part of the challenge posed by this scenario stems from uncertainty about non-moral matters, such as the efficacy of campaigns for political change, or about the exact circumstances of the person asking for help. But at least some part of what makes this choice hard is doubt of a purely moral nature. Doubt about the need for compassion towards those around us, versus the importance of efficient means to tackle structural injustice. Or doubt about how our interpersonal relationships matter, when compared to impersonal efforts to alleviate suffering, whoever it may befall. Even if you knew all of the non-moral facts of the situation, you might remain unsure of what to do because of this kind of moral doubt. One response to the moral doubt we face in choices such as this is to engage in reasoning and deliberation about how we ought to act; a concern for getting these choices right provides motivation and justification for the endeavour of moral philosophy. These sorts of situations are worth thinking about, in order to arrive at stable and coherent moral views that might guide our actions.

However, much of the time, we must make choices when we are yet to settle the relevant moral questions and our considered moral judgements do not provide an unambiguous recommendation for how to act. Compelling moral arguments often pull in opposing directions and even our best efforts often fall short of decisively settling the matter at hand. In the scenario described above, you will have to decide how to act in the here and now, even if you remain uncertain about the right course of action to take. The four self-contained papers in this thesis are united by an interest in understanding choices made in

this state of moral doubt. They are ultimately concerned with how we ought to reason when faced with choices in which our considered moral judgements provide conflicting recommendations, leaving us unsure of how to act.

An important idea running through these papers is the observation that how one responds to moral doubt depends, crucially, on the nature of one's moral judgements. For example, if one takes our moral attitudes to be, in all important ways, equivalent to beliefs about non-moral matters of fact, then there appears no reason to treat moral doubt any differently from non-moral uncertainty. If the moral judgements that pertain to the choice of how to help the homeless are equivalent to one's beliefs about what the weather holds tomorrow, then we ought to respond to uncertainty in the former in the same manner as uncertainty in the latter. Conventionally, this would involve modelling the uncertainty with probabilistic credences and modelling the choice with expected utility theory. If, on the other hand, our moral judgements are somehow different from our beliefs about non-moral matters, such the weather or which way up a tossed coin will land, then we may require a different way to understand moral doubt.

Thus, how we handle doubt and conflict in our moral judgements may partly come down to questions of meta-ethics. Another way of seeing the link between moral doubt and meta-ethics, is that one's views on the former may reveal or inform one's views on the latter. For example, if one thinks that maximising expected utility is the right way to make choices under empirical uncertainty, but think that this adapting approach to choices under moral uncertainty is somehow wrong-headed, then there must be some relevant differences between the empirical beliefs and moral judgements. Elucidating what seems mistaken about a decision theoretic approach to choice under moral uncertainty provides a way of thinking about the nature of moral judgements and what sets them apart from non-moral beliefs.

There is one major dividing line in the field of meta-ethics that also divides this thesis: that between cognitivism and non-cognitivism regarding moral judgements. According to the former, our moral judgements take the form of beliefs. According to the latter, our moral judgements are better thought of as a conative, desire-like attitude instead. My aim here is not to argue for a particular view in the debate between cognitivism and non-cognitivism. Rather, I intend to show that moral uncertainty poses a challenge whether one's moral

judgements take the form of beliefs *or* desire-like attitudes, and to describe the ways in which moral doubt might be accommodated in our choices on either view. Indeed, there is some reason to believe that there are individual differences between people's moral psychology and therefore, whatever the merits and flaws of these competing meta-ethical views, it is worth thinking about how best to understand and respond to moral doubt, whatever form one's moral judgements take. [1]

Thus, the four standalone papers that comprise this thesis can be divided into two pairs: Chapters 1 and 2 take a cognitivist view, to analyse moral doubt in terms of degrees of belief in moral propositions: Chapter 1 argues that this form of moral uncertainty is relevant to how we ought to choose, given the subjective norms of instrumental rationality; and Chapter 2 proposes a particular way of accommodating moral uncertainty within a decision theoretic model of rational choice. Chapters 3 and 4 start from a non-cognitivist perspective, presenting an alternative understanding of moral doubt and its implications for rational choice: Chapter 3 demonstrates how non-cognitivism can accommodate moral doubt in terms of ambivalence in one's moral attitudes; and Chapter 4 draws out some implications of ambivalence (moral or otherwise), by exploring the way in which the overall weight of reasons is reflected in the preferences of rational agents. The remainder of this introduction presents a more detailed summary of each of these chapters.

## Chapter 1: *The Relevance of Belief: Subjective Norms Under Empirical and Moral Uncertainty*

This paper concerns the way in which our beliefs about what we ought to do affect what we ought to do. When making choices under empirical uncertainty, it is generally accepted that the norms governing how we ought to act are sensitive to our empirical beliefs. That is to say, they are subjective norms. It is only subjective norms that can provide action-guidance for uncertain agents and make sense of the claim that we ought not to risk harming others, even if no harm in facts comes about. However, when the uncertainty

---

[1] Michael B. Gill, "Metaethical Variability, Incoherence, and Error," in *Moral Psychology*, ed. Walter Sinnott-Armstrong (Cambridge: MIT Press, 2008), 387–402; Walter Sinnott-Armstrong, "Mixed Up Meta-Ethics," *Philosophical Issues* 19, no. 1 (2009): 235–56; Jennifer Cole Wright, "The Fact and Function of Meta-Ethical Pluralism: Exploring the Evidence," in *Oxford Studies in Experimental Philosophy, Volume 2*, ed. Tania Lombrozo, Joshua Knobe, and Shaun Nichols (Oxford: Oxford University Press, 2018), 119–50.

concerns moral, rather than empirical matters, things are not so clear. When considering the relevance of our moral beliefs, there is disagreement about whether the appropriate norms are subjective, such that we simply ought to do whatever we believe we ought to do, or objective, such that our moral beliefs are completely irrelevant to the moral status of our actions. This presents a dilemma, because subjectivism regarding moral belief entails that an agent cannot have done something wrong, as long as they have done what they believe is right, while objectivism fails to be action-guiding under moral uncertainty. In this paper, I show that this dilemma is not as troublesome as it may first appear. I argue (i) that whether or not one's beliefs are relevant to what one ought to do may vary according to the context, and (ii) that there is an indirect role for morality in action-guidance under moral uncertainty, via agents' moral beliefs and the subjective norms of instrumental rationality. This demonstrates the need to find the right norms of rational choice under moral uncertainty.

## Chapter 2: *Preference and Commitment Under Moral Uncertainty*

This paper provides an answer to the question of what agents rationally ought to do when they are uncertain about what they morally ought to do. The Analogical View states that we should treat moral uncertainty as analogous to empirical uncertainty in models of rational choice. Existing instantiations of this view, such as Maximisation of Expected Choiceworthiness, face the problems of (i) severely limited scope, (ii) intertheoretic comparisons and (iii) swamping by 'fanatical' theories. In order to provide a unifying diagnosis of these problems, I show that these attempts to realise the Analogical View involve a crucial mistake: they introduce values into an agent's decision calculus that represent moral theories themselves, rather than the desirability of options to the agent, on the hypothesis that a given moral theory is true. I show that we can avoid these problems, while preserving the advantageous features of the Analogical View, by assigning values to options under a given moral theory by asking something like "*if this theory were true, how much would I desire this option?*" rather than "*if this theory were true, how much value would this theory assign to this option?*". Moreover, this alternative method of value assignment provides a role for the preferences, desires, or goals of rational agents that is curiously absent from the existing discussion of what individuals rationally ought to do when they are uncertain about what they morally ought to do.

# Chapter 3: *Attitudinal Ambivalence: Moral Uncertainty for Non-cognitivists*

This paper takes an alternative approach to the two that precede it. While much existing literature analyses moral doubt in terms of uncertainty in one's moral beliefs, non-cognitivists argue that moral judgements express a kind of conative attitude, more akin to desire than belief. Here, I present a scientifically informed reconciliation of non-cognitivism and moral doubt. The central claim is that attitudinal ambivalence can play the role of moral doubt for non-cognitivists. On this view, moral doubt is not uncertainty in a person's beliefs, but is ambivalence in their non-cognitive attitudes—that is, the presence of conflicting reasons that arise from these moral attitudes and support mutually exclusive options. I demonstrate that ambivalence has all of the features that we would expect it to have in order to play the role of moral doubt: it is gradable, can vary through time, covaries with strength of motivation, and is suitably distinct from the other structural features of our moral judgments. The principal upshot is that one can consistently be a non-cognitivist and recognise variation in the certitude of moral judgments. However, the differences between ambivalence and uncertainty also suggest a different understanding of, and practical response to, the phenomenon of moral doubt. This poses a new challenge: what should we do when ambivalent about what to do? This is a challenge that must be met by cognitivists and non-cognitivists alike.

# Chapter 4: *The Balance and Weight of Reasons*

The final paper in this thesis explores the implications of ambivalence for rational agency, by providing a detailed characterisation of some ways in which our preferences reflect our reasons. I demonstrate that practical reasons can be characterised along two dimension that influence our preferences: their balance and their weight. This is analogous to a similar characterisation of the way in which probabilities reflect the balance and weight of evidence in formal epistemology. Much attention in the philosophical literature related to practical rationality concerns the balance of reasons. That is to say, the degree to which our reasons favour one option over another, or not, as the case may be. However,

comparatively little attention has been paid to the role of the overall weight of reasons in our practical deliberation. Here, I begin to fill this gap, showing how a full understanding of the weight of reasons is crucial for an adequate account of preference and choice. In particular, I demonstrate that the greater the weight of reasons underpinning an agent's evaluation of options in a choice, the more stable that evaluation will be in the face of new reasons. The upshot is a more complete picture of a particular kind of decision, labelled by Isaac Levi and, more recently, Ruth Chang as "*hard choices*". These are choices in which one option is better than another in some ways, the other is better than the first in some ways, but neither seems better overall. The distinction between the balance and weight of reasons provides a new way of understanding how hard choices vary by degree and explains what it is that makes them so hard.

# CHAPTER 1

# The Relevance of Belief: Subjective Norms Under Empirical and Moral Uncertainty

**Jiminy Cricket**: *When you get in trouble and you don't know right from wrong,*
*Give a little whistle, give a little whistle.*
*When you meet temptation and the urge is very strong,*
*Give a little whistle, give a little whistle,*
*And always let your conscience be your guide.*

*Pinocchio* (1940)[2]

**Da Mayor**: *Doctor, this is da Mayor talkin'.*
**Mookie**: *OK. OK.*
**Da Mayor**: *Doctor, always do the right thing.*
**Mookie**: *That's it?*
**Da Mayor**: *That's it.*

*Do The Right Thing* (1989)[3]

## 1    Introduction

Do an agent's beliefs govern what they morally ought to do? If so, which are the relevant sorts of belief and what role does this leave for the facts? This paper is intended to address these questions. It concerns the ways in which what an agent ought to do depends on their beliefs about what they ought to do. This, in turn, depends on whether the norms governing choice under uncertainty are subjective or objective.

---

[2] Walt Disney, Ben Sharpsteen, and Hamilton Luske, *Pinocchio* (RKO Radio Pictures, 1940).
[3] Spike Lee, *Do The Right Thing* (Universal Pictures, 1989).

A subjective norm is one according to which what an agent ought to do depends upon their beliefs. For example, the injunction to "*let your conscience be your guide*" suggests that you ought to do what you believe you ought to do. What you ought to do depends on what you believe, so this is a subjective norm. Conversely, an objective norm does not relativise what you ought to do to what you believe you ought to do in this way. For example, what you ought to do according to the norm "*do the right thing*" does not depend on what you believe the right thing to be, so this is an objective norm. The term "subjectivism" shall here refer to the view that there are subjective norms above and beyond the objective norms that would apply if one had full knowledge of their situation. The term "objectivism" shall refer to the denial of these subjective norms. These positions are referred to variously in related literature as uncertaintism, actualism, probabilism, externalism and internalism, but I shall refer only to objectivism and subjectivism here. For simplicity, this discussion will proceed as if the true objective norms are a form of consequentialist ethics. I will talk of consequences rather than, say, virtues or rights, but the analysis herein could be adapted to apply to other classes of moral theory.

In §2, I will review the classic debate between subjective and objective conceptions of morality, motivating the view that there are subjective moral norms under empirical uncertainty. In other words, what agents ought to do depends on their non-moral beliefs about the world. In §3, I will consider the further claim that there are subjective moral norms that are sensitive to our *moral* beliefs. That is to say, what one ought to do depends on what one believes one ought to do. This possibility appears to raise a dilemma: if one ought to do what one believes one ought to do, then false moral beliefs seem to excuse all wrongdoing. If, on the other hand, what one ought to do is insensitive to one's moral beliefs, then morality fails to provide action-guidance under moral uncertainty. The remaining sections demonstrate that this dilemma is not as troublesome as it may seem at first. In §4 I show that whether or not moral ignorance excuses wrongdoing may vary from case to case and in §5 I show how moral norms can still play a suitable role in guiding action, via agents' moral beliefs and the subjective norms of instrumental rationality. The upshot is that we can seek norms of rational choice under moral uncertainty, without being committed to the view that false moral beliefs always excuse wrongdoing or that morality is not action-guiding.

## 2      Subjectivism Regarding Empirical Beliefs

Consider the following scenario:

> *Rosa is trying to decide whether or not to take her umbrella out with her, but she is uncertain about whether or not it will rain today. She believes that it is slightly more likely that it will not rain than that it will, but she cannot be sure. If it does not rain, it would be slightly better for her not to take her umbrella. If it rains, it would be much better for her to take her umbrella. In fact, it will not rain.*

Should Rosa take her umbrella or not? It seems as though there is intuitive support for answering both "yes" and "no" to this question. Given that it will not rain, Rosa will be better off if she does not have her umbrella with her. Therefore, Rosa should not take her umbrella. On the other hand, Rosa does not know that it will not rain. Given her beliefs, she would be taking a comparatively large risk by not taking her umbrella, for relatively little potential gain. Therefore, Rosa should take her umbrella.

Relative to what she believes, Rosa ought to take her umbrella. This is a subjective norm; it depends on Rosa's beliefs. Relative to the true state of the world, Rosa ought not to take her umbrella. This is an objective norm; it does not depend on Rosa's beliefs.

It is a commonly held view that subjective norms govern how rational agents should act. [4] Indeed, the central maxim of expected utility theory, the paradigmatic procedure for decision making under uncertainty, is a generalised subjective norm: maximise expected utility. In Rosa's case, expected utility theory would recommend that she take the umbrella if the expected value of taking the umbrella, given the chance of rain and the value to Rosa of the various possible outcomes, is higher than the expected value of not taking it. This should not be interpreted as prioritising expected utility over actual utility, nor of claiming that one should adhere to subjective, but not objective, norms. If Rosa knew which of the available options would in fact provide the greatest utility, she should take that option. It

---

[4] Although for the suggestion that rationality may be actualistic, rather than probabilistic, see: Michael A Slote, *Beyond Optimizing: A Study of Rational Choice* (Cambridge: Harvard University Press, 1989).

is only when this information is not available that she must fall back on the subjective norm of maximising expected utility. This is a norm of rationality; to do otherwise would be irrational. Note that, when one knows the outcome of an action with certainty, maximising expected utility is exactly equivalent to maximising actual utility.

In this example, the only value at stake is prudential. Rosa has only herself to worry about and there are no morally significant differences between her options. Now consider the following example, structurally similar to Rosa's decision, but with a salient moral dimension:

> *Shawn is trying to decide whether or not to administer a pain-relieving medicine to his mother.*
> *He is uncertain about whether or not the medicine will interact with another of his mother's*
> *medicines and kill her. He believes that it is slightly less likely that it will kill his mother than*
> *that it will not, but he cannot be sure. If there is no drug interaction, it would be slightly better*
> *for his mother to have the pain killer. If there is a drug interaction, she will die. In fact, there*
> *will be no deadly drug interaction.*

Should Shawn give his mother the medicine or not? As with Rosa's decision about the umbrella, there appear to be conflicting objective and subjective norms dictating what Shawn ought to do. Objectively—relative to the facts of the scenario—he should give his mother the drug, as this will in fact have the best consequences. She will be made better off by the pain relief and will not be harmed. But subjectively—given Shawn's beliefs about the scenario—it seems as though he ought not to give his mother the drug. It would be reckless to risk his mother's life for the sake of some slight pain relief and it seems clear that this would be morally wrong.

Questions about subjective norms arise when a decision maker has incomplete information about their situation. If Shawn knew that the medicine would not kill his mother, it would be clear that he ought to administer it. However, because he lacks complete information about his mother's medicines, he is uncertain about what he ought to do. This may lead

him to ask, "*What ought I to do when I do not know what I ought to do?*" [5] Despite the similarities between Rosa's and Shawn's situations, cases like these are not always interpreted as equivalent and there is disagreement among philosophers about how to answer this question. Objectivists about morality may well agree that agents ought to maximise their own expected utility when faced with non-moral decisions, but they deny the existence of subjective norms for moral decisions. They claim that it does not matter whether or not one knows what one ought to do, there is simply an objective fact of the matter about how an agent ought to act and nothing more. There are no additional subjective norms that govern how agents should act relative to their beliefs about some particular features of their situation. On this view, what such an agent ought to do is exactly what they ought to do if they did have all of the relevant information. [6]

This view might appear to be attractive in its parsimony and the weight it gives to the facts of the matter in determining how one ought to act. It also gives a straightforward account of praise and blame that only requires analysis of the actual events, rather than any information about an agent's beliefs. Simply put, an agent is to blame for their actions if they do not act as they objectively ought. It may have further appeal to those who hold the intuition that what matters most, morally speaking, is what actually happens, rather than what anybody believes or intends.

However, there are significant problems with this form of moral objectivism. Firstly, its ascription of right and wrong conflicts with our considered moral judgements in cases of pure risk—situations in which a course of action risks harming a person, but this harm does not in fact come to pass.[7] For example, suppose that one person was to hold a gun to the head of another and play Russian Roulette. Suppose that no bullet ends up firing and the victim never finds out what has happened, such that they never actually come to any harm. Nonetheless, it seems hard to deny that it would be severely wrong to put

someone at such risk of harm. For this reckless act to be wrong, it must be so subjectively, as the risk only exists relative to the agent's lack of information about the location of the bullet. An objective account of morality cannot accommodate this judgement, as it cannot relativise the moral status of the action to the agent's beliefs, but this is exactly what determines its wrongness. To play Russian Roulette in this case is wrong precisely because the agent does not know the fact that no harm will result. Thus, objectivism is incompatible with the judgement that exposing others to pure risks can be morally wrong.

Secondly, objectivism has a practical limitation that has led many philosophers to reject it. Under uncertainty, it fails to perform the primary role of any normative theory: providing a guide to action. [8] Recall the case of Shawn. He does not know what he objectively ought to do, because he does not know which action will have the best consequences. To advise Shawn that what he ought to do is simply that which will have the best consequences will be no help in guiding his action. This is precisely what he does not know. Yet objectivism is unable to offer anything more. To the question of "*What ought I to do when I do not know what I ought to do?*" the objectivist can only answer "*Do what you would do if you did know what you ought to do.*" But this is an impossible instruction to follow! Thus, objectivism is not action-guiding under uncertainty. Note that the fact that objectivism is not action-guiding does not render it entirely useless. Moral principles are used to appraise, as well as to guide action. Even if objective norms are unable to guide action, they might still give an account of which actions are right or wrong, or which are worthy of praise and which of blame. We will return to this point below.

So, in order to guide action and to make sense of the wrongness of pure risk imposition, we must invoke subjective moral norms. When one does not know how one ought objectively to act, then one ought, subjectively, to act as these norms dictate. [9] Many subjectivists take the correct subjective norm for moral decision making under uncertainty to be maximisation of expected value, in much the same way as for self-interested rational

---

[8] Frank Jackson, "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection," *Ethics* 101, no. 3 (1991): 461–82; Zimmerman, "Is Moral Obligation Objective or Subjective?"; William MacAskill and Toby Ord, "Why Maximize Expected Choice-Worthiness?," *Noûs* 54, no. 2 (2020): 327–53.
[9] For more on uncertainty, action-guidance and subjectivism (concerning reasons, rather than norms), see: Andrew Sepielli, "Subjective and Objective Reasons," in *The Oxford Handbook of Reasons and Normativity*, ed. Daniel Star (Oxford: Oxford University Press, 2018), 784–99.

choice. [10] While Rosa should choose the option with the highest expected value to her, Shawn should aim to maximise expected *moral* value. This does not specify what moral value itself consists in, but is consistent with a range of conceptions of objective morality. There are other possible subjective norms, such as a satisficing account of subjective consequentialism, [11] a rule prescribing the action that is most likely to have the best actual consequences, [12] or non-consequentialist approaches concerning agents' intentions [13] or the quality of their will. [14] A comparison of these subjective norms is not within the scope of this paper; the point here is simply that an analysis of moral subjectivism need not be an analysis of decision theoretic consequentialism. However, this issue is complicated by consideration of a particular variety of uncertainty under which we must make decisions: moral uncertainty.

## 3      Subjectivism Regarding Moral Beliefs

Consider another decision problem:

> *Alex is trying to decide whether to buy a leather bag or a faux leather alternative. She is uncertain about the moral value of animal lives. She believes that it is slightly less likely that animal lives have moral value than that they do not, but she cannot be sure. If animal lives do not have moral value, it would be slightly better for her to buy the leather bag. If animal lives do have moral value, then it would be severely morally wrong to buy the leather bag. In fact, animal lives have no moral value.*

---

[10] For example: Derek Parfit, *Reasons and Persons* (Oxford University Press, 1984); John Broome, *Weighing Goods: Equality, Uncertainty and Time* (Wiley-Blackwell, 1991); Jackson, "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection."

[11] Michael Slote and Philip Pettit, "Satisficing Consequentialism," *Proceedings of the Aristotelian Society* 58 (1984): 139–63.

[12] A view that Jackson (1991) attributes to Peter Railton (1984).

[13] Philippa Foot, "The Problem of Abortion and the Doctrine of Double Effect," *Oxford Review* 5 (1967): 5–15; Gertrude Elizabeth Margaret Anscombe, "Medalist's Address: Action, Intention and 'Double Effect,'" *Proceedings of the American Catholic Philosophical Association* 56 (1982): 12–25, https://doi.org/acpaproc19825611.

[14] Peter Strawson, "Freedom and Resentment," *Proceedings of the British Academy* 48 (1962): 25, https://doi.org/10.1073/PNAS.48.1.1.

Alex, like Shawn, is faced with a decision about what she morally ought to do. As in Shawn's decision under empirical uncertainty, there seems to be a difference between what Alex ought to do according to what is true, and what she ought to do relative to her beliefs. Objectively, she should buy the leather bag, because it will be better for her and, *ex hypothesi*, animal lives have no moral value. Subjectively, however, she should not, because to do so would be reckless given her belief that it might be severely morally wrong to buy leather.

However, there is a crucial difference between these decisions: whereas Shawn is uncertain about the consequences of his decision, Alex is uncertain about a purely moral matter. Shawn may be entirely certain of which moral statements are true, but remain unsure of what to do, because he does not know all of the relevant non-moral facts. This is empirical uncertainty. Alex, on the other hand, may be entirely certain of the consequences her action will have, but remain unsure of what to do, because she does not know which moral statements are true. This is moral uncertainty.

Drawing a distinction between moral and empirical uncertainty follows the increasingly standard terminology in the philosophical literature on this topic. It should be noted, however, that I do not mean for these terms to imply anything in particular about the nature of moral judgements themselves. The contrast with 'empirical' beliefs should not be taken to be in tension with any meta-ethical view according to which moral matters are themselves empirical, in some sense. On this view, there may be no distinction to be drawn between moral and empirical beliefs. Nor should it imply that all beliefs about non-moral matters are of the same kind and should be handled in the same way in our decisions. Whatever terms we use, the key distinction I have in mind here is that between beliefs about propositions that express moral judgements and beliefs about those that do not.

How to make decisions under moral uncertainty has thrown up thorny issues above and beyond those relating to ordinary empirical uncertainty. There is a growing discussion concerned with identifying the correct principles for choice under moral uncertainty, much of which has focussed on whether or not it is possible to accommodate moral uncertainty

in decision theoretic models of rational choice. [15] However, this presupposes that we should take moral uncertainty into account in the first place. For a decision to be sensitive to an agent's uncertainty, moral or otherwise, there must be subjective norms that govern how one ought to act relative to the beliefs that comprise this uncertainty. Any proposal for how to factor moral uncertainty into decision-making will be inherently subjectivist. On the face of it, this appears to lead us to the very question covered in the previous section: are there any subjective moral norms? What Alex should do depends on whether or not there exist subjective norms, just as it did in the previous cases of Rosa and Shawn. However, one may consider this an altogether different question when it concerns moral, rather than empirical, beliefs.

Brian Weatherson and Elizabeth Harman have both claimed that how one ought to act depends on one's non-moral beliefs, but not on one's moral beliefs.[16] In other words, they are moral subjectivists about decisions under empirical uncertainty, but moral objectivists about decisions under moral uncertainty. According to this view, there are no additional subjective norms to govern our choices under moral uncertainty; how one should act is determined by the true moral theory, which ascribes rightness and wrongness relative to an agent's beliefs about the non-moral features of their actions, but not relative to their beliefs about morality. Harman argues for this position by claiming that subjectivism under moral uncertainty entails an implausible conclusion: that holding a false moral belief exculpates an agent from blame for wrongdoing. Contra subjectivism, agents are blameworthy if they act wrongly, even if they do what they believe they ought to do, so subjectivism must therefore be false. The view that false moral beliefs do not exculpate agents from blame is supported by cases in which it would seem clear that a person is to blame for a wrongful act, even if they do not believe that their act is wrong. Indeed, many such cases are paradigmatic examples of blameworthy acts.

Harman asks us to imagine, for example, a gang member who believes that they have a strong moral obligation to maintain loyalty to the gang and to protect its interests. Suppose

---

[15] See Chapter 2.

[16] Brian Weatherson, "Running Risks Morally," *Philosophical Studies* 167, no. 1 (2014): 141–63; Elizabeth Harman, "The Irrelevance of Moral Uncertainty," in *Oxford Studies in Metaethics, Volume 10*, ed. Russ Shafer-Landau, vol. 10 (Oxford University Press Oxford, 2015), 53–79.

that this loyalty requires harming an innocent person in the interests of the gang. The gang member gives a high credence to the view that they are morally obliged to harm this person and that it would be seriously wrong not to. If subjective moral norms make it right to act in accordance with one's moral beliefs, then it would be right for the gang member to carry out this harmful act. But this must be incorrect; hurting innocent people in the interests of a gang is clearly a grave and blameworthy moral wrong. The gang member's false moral beliefs cannot exculpate them from blame for their wrongful act. Thus, claims Harman, blameworthiness must be determined by the actual moral theory, not one's beliefs about morality.

The view that agents are not to blame for wrongdoings that result from their own false empirical beliefs, but are to blame for wrongdoings that result from their false moral beliefs is not entirely new. This idea echoes the following two passages from Aristotle:

> "*individuals and legislators punish and take vengeance on those who do wicked acts (unless they have acted under compulsion or as a result of ignorance for which they are not themselves responsible)*" [17]

> "*we punish those who are ignorant of anything in the laws that they ought to know and that is not difficult, and so too in the case of anything else that they are thought to be ignorant of through carelessness; we assume that it is in their power not to be ignorant, since they have the power of taking care.*" [18]

The first passage suggests that an agent is not to blame when their actions result from ignorance, as long as they are not themselves responsible for that ignorance. The second passage suggests that agents *are* to blame when their action results from ignorance of "*the laws*", since it is in their power not to be so ignorant.

In fact, we needn't even enter into a discussion of blame in order to see this point. The philosophical literature on blame is extensive and complex, but on most accounts, for an agent to be blameworthy they must have done something wrong and meet some further

---

[17] *Nicomachean Ethics*, III.5, 1113b.
[18] *Nicomachean Ethics*, III.5, 1113b-1114a.

criteria, such as having or lacking certain intentions, beliefs, or abilities. But if we invoke subjective moral norms that tie what an agent ought to do to what they believe they ought to do, then we cannot even judge them to have done something wrong when acting as they believe they ought, regardless of the further question concerning whether or not they meet the conditions for blameworthiness. [19] It is only objectivist views, exemplified by Harman's "*actualism*", that are able to provide an appropriate account of right and wrong, in line with intuitions about how to judge the actions of those who hold false moral beliefs. Note that the example of the gang member is not only a toy philosophical case, designed to motivate an abstract point. There is a growing consensus in the fields of sociology and anthropology that many, perhaps even most, acts of violence are principally motivated by strong moral convictions on the part of the agent. [20] If we are to uphold the judgement that many such acts are morally wrong, it follows that moral ignorance does not absolve wrongdoers from responsibility and, therefore, that moral subjectivism must be mistaken.

However, objectivism under moral uncertainty faces exactly the same difficulty that led us to abandon objectivism under empirical uncertainty. Unless a moral system includes subjective norms that are sensitive to moral beliefs, it is of little use to the individual facing a decision while uncertain of the right answers to the relevant moral questions. In other words, objectivism under moral uncertainty fails to be action-guiding for the morally uncertain agent. Thus, we appear to be on the horns of a dilemma: either moral ignorance always excuses wrongdoing, or morality fails to be action-guiding. In what follows, however, I will describe an alternative view that avoids the worst of these two conclusions.

## 4    Ignorance and Responsibility

Firstly, we should avoid jumping too quickly from the observation that there are *some* cases in which moral ignorance does not absolve agents from responsibility, to the conclusion that an agent's moral beliefs are *never* relevant to how they should act. It does seem mistaken to say that Harman's gang member did the right thing, simply because they did

---

[19] I thank Chloé de Canson for highlighting this point in personal communication.
[20] Elijah Anderson, *Code of the Street: Decency, Violence, and the Moral Life of the Inner City* (WW Norton & Company, 2000); Alan Page Fiske and Tage Shakti Rai, *Virtuous Violence: Hurting and Killing to Create, Sustain, End, and Honor Social Relationships* (Cambridge: Cambridge University Press, 2015).

what they believed they morally ought to do. Nevertheless, there is some sense in which what they did was understandable, given their moral beliefs. Perhaps even admirable. To inflict unwarranted harm on a person in the interests of the gang is wrong, but to do so on the basis of a strong moral conviction at least shows a desire to do the right thing, along with a willingness to make personal sacrifice to serve this end. We could think of a different agent, who carries out the same course of action knowing full well how morally wrong it is. This person seems to have done something worse than in the original example. They have also harmed an innocent person, but they have done so with an indifference towards their moral beliefs and a disregard for the importance of doing the right thing.

There are at least three things that we tend to think moral agents should get right: (1) they should have the correct moral beliefs, (2) they should act in a morally conscientious way and (3) they should act in the objectively right way. (1) and (2) in combination guarantee (3). When agents come to the correct moral beliefs and act in accordance with them to do the right thing, they are plainly acting as they ought. When agents achieve some, but not all of (1)-(3), however, matters are more complicated. Harman's gang member, for example, does what they believe they ought to do, so achieves (2), but neither holds true moral beliefs, nor objectively does the right thing, so fails to achieve (1) and (3). They appear to have done something less wrong than the alternative agent who holds correct moral views (1), but fails to do the objectively right thing (3) because they do not act as they believe they ought (2). There are also cases in which someone might come to hold a false moral belief, but end up doing the right thing nonetheless, by failing to do what they believe they ought to do.[21]

The all-things-considered judgements we come to in these sorts of cases depend on the precise ways in which we balance out the features of agents' beliefs, motivations, and actions. This weighing up of different considerations must also take into consideration how agents arrive at their belief and motives in the first place, as it would be a mistake to overlook the factors bring about a particular agent's moral psychology. These factors include contingent circumstances, such as opportunities for moral education and

---

[21] The well-known case of Huckleberry Finn and Jim is an example of this sort of scenario. See Zoë Johnson King, "Accidentally Doing the Right Thing," *Philosophy and Phenomenological Research* 100, no. 1 (2020): 186–206, for a recent discussion of this topic.

reflection, or the availability of role models and confidants, as well as the role that people play in cultivating their own moral character. For example, if a person whom we judged to have done something wrong did so because they were indoctrinated since childhood, radicalised, or brainwashed, we should have some sympathy for them and view their actions with an appreciation of their own experience and the broader context. In some sense, such individuals are also victims of the wrong they carry out, since their ability to do otherwise is undermined by those with power over them. In contrast, if someone has all of the opportunities in life to think freely about their rights and duties, to read moral arguments and talk to informed and well-intentioned role models, then they seem to be entirely responsible if they end up making rash and impulsive moral judgements that lead to false beliefs and immoral actions. They bear greater responsibility for failing to arrive at the right moral beliefs and for the immoral acts that they consequently carry out. The person who has been nurtured and encouraged to reflect on their moral duties in a safe space is simply not starting from the same point as someone brought up in a world of violence and oppression, so their future choices should not be evaluated on the same terms. The culpability of an agent for some action that is motivated by false moral beliefs depends, *inter alia*, on how responsible that agent is for forming that false belief in the first place. Again, this point echoes Aristotle:

> "*we punish a man for his very ignorance, if he is thought responsible for the ignorance, as when penalties are doubled in the case of drunkenness; for the moving principle is in the man himself, since he had the power of not getting drunk and his getting drunk was the cause of his ignorance.*" [22]

The discussion of precisely when an agent is responsible for their own moral views and, hence, their actions and choices raises a whole battery of thorny and complicated issues that cannot be untangled in one go. How to evaluate cases of accidentally doing the right thing and doing the wrong thing for the right reasons requires a detailed and nuanced account that I do not aim to provide in this paper. But we do not need answers to all of these questions in order to grasp the point that matters for present purposes—that it would be too simplistic to think that there is a neat dichotomy between full subjectivism on the one hand, according to which what one ought to do is entirely determined by one's moral

---

[22] *Nicomachean Ethics*, III.5, 1113b.

beliefs, and full objectivism on the other, according to which one's moral beliefs are utterly irrelevant. Harman's argument is something like the following: subjectivism entails that moral ignorance is exculpatory, moral ignorance is not exculpatory, so subjectivism must be false. The considerations here, meanwhile, suggest a more moderate alternative: subjectivism entails that a person does not do something wrong, as long as they act in accordance with their moral beliefs; there are cases in which a person does in fact do something wrong, even though they act in accordance with their beliefs; therefore, there are circumstances in which subjectivism does not hold. Sometimes agents ought to act in accordance with their moral beliefs, but sometimes we might judge an agent's actions to be wrong even if they do what they believe is right. What matters here is whether or not the relevant norms are acceptable in and of themselves and, as we have seen, there are at least some scenarios in which we might plausibly endorse moral norms that do not absolve morally ignorant agents of responsibility in the way that subjective norms would. Therefore, there are some cases of moral ignorance or uncertainty in which there are only objective moral norms and some in which there are subjective moral norms.

## 5 Morality, Rationality, and Action-guidance

Even this watered-down version of Harman's objectivist view faces a substantial challenge. As soon as we stray from the fully subjectivist position that you always ought to do whatever you believe you ought to do, morality ceases to be action-guiding under moral uncertainty. As long as there are some cases in which there are no subjective norms, there are some cases in which we lack action-guidance. However, the question of whether morality itself provides subjective norms does not necessarily settle the question of whether *any* subjective norms exist to aid decision making under moral uncertainty. Moral reasoning is difficult, so it is to be expected that agents will remain uncertain about at least some moral problems. Nonetheless, we are frequently forced to make moral decisions while this uncertainty persists and we need some way of establishing how we "ought" to proceed, even if this "ought" is not what some would consider to be a part of morality.

Recall the decision facing Alex, described above. Suppose for argument's sake that the objectivists are right and that all moral theory can tell us is that animal lives have no moral

value, so Alex ought to buy the leather bag. Alex does not know this fact, but she must make a decision nonetheless and will have to find some way to do so, given her uncertain moral beliefs. There are many possible ways of making this decision and they do not all seem equally appropriate. She could flip a coin. She could choose the option that she believes is most likely to have the morally worst outcome. She could choose the option with the highest expected value. Clearly some ways of making a decision like this are better, in some sense, than others. The first option seems arbitrary and involves needlessly ignoring the beliefs she does have, simply because she is not fully informed. The second seems actively malevolent. The last option at least coheres with her beliefs and with some common intuitions about situations like this. Clearly it is possible to debate the merits and flaws of possible approaches to choice under moral uncertainty, but this immediately raises a further question: if morality does not provide the norms for choice under moral uncertainty, in what sense are some decision procedures better or more appropriate than others?

An answer presents itself if we return to the first case considered in this paper, Rosa and her umbrella. Recall that the subjective norm that recommended taking the umbrella—maximisation of expected utility—is a norm of rationality. Given Rosa's desires and beliefs, it would be rational for her to take the umbrella even though it would be objectively better for her not to. Similarly, in cases of moral uncertainty we can look to principles of rationality to guide our choices in light of our particular beliefs and desires. This approach has been adopted [23] and explicitly defended [24] by numerous writers on the subject of moral uncertainty.

In discussing what one *rationally* ought to do when uncertain about what one *morally* ought to do, we need not be committed to any particular moral view. Therefore, we can consider the rational basis for different approaches to choice under moral uncertainty without

---

[23] Jacob Ross, "Rejecting Ethical Deflationism," *Ethics* 116, no. 4 (2006): 742–68; Ted Lockhart, *Moral Uncertainty and Its Consequences* (Oxford: Oxford University Press, 2000); William MacAskill, Krister Bykvist, and Toby Ord, *Moral Uncertainty* (Oxford: Oxford University Press, 2020).

[24] Krister Bykvist, "Evaluative Uncertainty, Environmental Ethics, and Consequentialism," in *Consequentialism and Environmental Ethics*, ed. Avram Hiller, Ramona Ilea, and Leonard Kahn (New York: Routledge, 2014), 122–35; Krister Bykvist, "Some Critical Comments on Zimmerman's Ignorance and Moral Obligation," *Journal of Moral Philosophy* 15 (2018): 383–400; Andrew Sepielli, "What to Do When You Don't Know What to Do When You Don't Know What to Do ...," *Noûs* 48, no. 3 (2014): 521–44.

having settled the questions raised in §3 about the ways in which one's moral beliefs affect what one morally ought to do. More precisely, we may speak of the *instrumental* rationality of an agent's choices. That is, the appropriateness of the means that they adopt in pursuit of their ends, whatever those ends may be. Conventional accounts of instrumental rationality take as their starting points the desires and beliefs of the agent in question. A choice is instrumentally rational insofar as it is the best way of satisfying an agent's desires, given their beliefs. Rosa's desires are to stay dry and to avoid unnecessarily carrying her umbrella. Given her belief about the chance of rain and the strength of her desire to stay dry, it is instrumentally rational for her to take the umbrella, even if the weather happens to stay fair. Likewise for Alex, her desires and beliefs determine what it would be rational for her to do, irrespective of what the relevant moral norms actually say. Given her moral motivation and her beliefs about the possible wrongness of buying leather, it is instrumentally rational for her to buy the faux leather alternative, even if this wouldn't in fact be morally required.

The distinction drawn here between morality and rationality will collapse if one is convinced by recent attempts to reduce the former to the latter by constitutivists such as Christine Korsgaard, [25] David Velleman, [26] and Michael Smith, [27] or, for that matter, Immanuel Kant before them. [28] By relying on this distinction, the proposal here is not compatible with such views on the relation between rational agency and morality. However, it is worth noting that this presents a challenge to these views themselves: as I have noted, there is a tension between the need for a view of morality that is action-guiding and one that avoids entailing that moral ignorance always excuses wrongdoing. Whether any constitutivist meta-ethical theory has the resources to resolve this tension remains an open question. For the time being, I will restrict the discussion on the assumption that morality and rationality are not connected in any of the ways that constitutivists propose. On this view, considerations of instrumental rationality can provide some grounds for guiding and evaluating an agent's choices, without being committed to saying anything in

---

[25] Christine Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996); Christine Korsgaard, *Self-Constitution: Agency, Identity, and Integrity* (Oxford: Oxford University Press, 2009).
[26] David Velleman, *Self to Self* (New York: Cambridge University Press, 2005); David Velleman, *How We Get Along* (Cambridge: Cambridge University Press, 2009).
[27] Michael Smith, "A Constitutivist Theory of Reasons: Its Promise and Parts," *Law, Ethics and Philosophy* 1 (2013): 1–30; Michael Smith, "The Magic of Constitutivism," *American Philosophical Quarterly* 52, no. 2 (2015): 187–200.
[28] Immanuel Kant, *Groundwork of the Metaphysics of Morals (1785)*, ed. Mary Gregor and Jens Timmerman (Cambridge: Cambridge University Press, 2012).

particular about their moral status. The question of whether or not acting in accordance with false moral beliefs absolves agents of responsibility simply has no bearing on the instrumental rationality of a choice, and vice versa.

Conceiving of choices under moral uncertainty as being governed by the norms of instrumental rationality is part of the reason that maximisation of expected choice-worthiness (MEC), an analogue of expected utility theory, has been proposed as the appropriate subjective norm for decisions under moral uncertainty. [29] Insofar as expected utility theory is rational under empirical uncertainty, MEC is said to be rational under moral uncertainty. [30] Lurking beneath this proposal, as well as in the analysis presented in this paper so far, is the tacit assumption that moral judgements are much like empirical beliefs and play the same role in rational choice. This assumption is just one possible view in a hotly contested meta-ethical battleground and there is a great deal of detail to be filled in about how moral values make it into an agent's decision calculus. Nonetheless, we may evaluate choices from the perspective of instrumental rationality, if we consider the particular desires and beliefs of the decision maker in question. There is ongoing debate about whether MEC or some alternative account provides the correct principles for choice under moral uncertainty, [31] but those in the opposing camps at least agree that the dispute is between different accounts of how it would be *rational* to act. [32] Whatever the correct subjective norms, they are norms of rationality.

Note that this does not mean that moral theory says nothing in cases of moral uncertainty, nor that it is trumped or outranked by rationality. [33] There will still be morally right and wrong options, no matter what one's beliefs about morality. It is just that the agent in

---

[29] MacAskill and Ord, "Why Maximize Expected Choice-Worthiness?"; MacAskill, Bykvist, and Ord, *Moral Uncertainty*.

[30] See Chapter 2 for an in-depth discussion of this proposal.

[31] See, for example: Lockhart, *Moral Uncertainty and Its Consequences*; Johan E Gustafsson and Olle Torpman, "In Defence of My Favourite Theory," *Pacific Philosophical Quarterly* 95, no. 2 (2014): 159–74; Edward J Gracely, "On the Noncomparability of Judgments Made by Different Ethical Theories," *Metaphilosophy* 27, no. 3 (1996): 327–32.

[32] Lockhart, *Moral Uncertainty and Its Consequences*; Ross, "Rejecting Ethical Deflationism"; Andrew Sepielli, "Normative Uncertainty for Non-Cognitivists," *Philosophical Studies* 160, no. 2 (2012): 191–207; Ittay Nissan-Rozen, "Against Moral Hedging," *Economics and Philosophy*, no. 3 (2015); MacAskill and Ord, "Why Maximize Expected Choice-Worthiness?"

[33] For an argument in favour of thinking about what is morally, rather than rationally, required under moral uncertainty, see: Chelsea Rosenthal, "What Decision Theory Can't Tell Us About Moral Uncertainty," *Philosophical Studies*, 2020, https://doi.org/https://doi.org/10.1007/s11098-020-01571-3.

question may not know the true moral norms, so will require an account of how it is rational to act given their particular moral beliefs and their desire to act morally. Similarly, it is not the case that morality trumps or outranks rationality when there is no moral uncertainty. There is simply no conflict between the two for a morally motivated rational agent.

By turning to the subjective norms of instrumental rationality, we can provide a principled guide to choice under moral uncertainty, while avoiding the implausible conclusion that acting in accordance with one's own moral views always excuses wrongdoing. However, even this may not appear to be adequate. One may think that it is not enough simply to have *some* principles by which to guide action, but that morality itself must be action-guiding. Actions and their consequences are the subject of many moral statements and we naturally look to moral principles when considering how to act; surely morality's raison d'etre is to guide behaviour. In the words of Frank Jackson, "*the passage to action is the very business of ethics.*" [34] However, there are a number of different claims we may be making when we say that a given moral theory is not action-guiding. It is important to clearly distinguish these different failures of action-guidance, so as not to mistake a commitment to the one form of action-guidance for an unrelated challenge to objectivism under moral uncertainty.

Perhaps most simply, we may be making a descriptive statement about whether or not some real person happens to be acting in accordance with a given moral theory. If a moral principle says "*Thou shalt not kill*", but one person takes the life of another, then this moral principle has failed to guide this person's action. Almost any moral principal worth stating has been violated at some point or other, so clearly most, if not all, moral theories fail to be action-guiding in this sense at least some of the time. Moreover, the very argument in favour of objectivism under moral uncertainty—that moral ignorance is not exculpatory—presupposes that people are sometimes ignorant of, and therefore not guided by, the true objective moral norms. Perhaps there is something to be said for moral views not to stray *too* far from real human behaviour. A moral theory that demanded a way of life that was unlike the behaviour of any real and morally conscientious person may

---

[34] Jackson, "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection," (p.467).

stretch credibility. But this must be because of the features of the theory itself. The fact that people don't act in accordance with a moral principle is not itself a reason to doubt that principle. For example, the fact that so many people in the United States used to support and enact a system of chattel slavery poses no challenge to the view that such practices are morally reprehensible. If a moral theory is highly implausible, it might not have many adherents, but the fact that no one acts in line with such a view is due to its implausibility, not the other way around. Indeed, one of the principal purposes of moral thought is to provide grounds for changing conduct that we deem to be inappropriate, and there must be some divergence between moral theory and existing behaviour in order for there to be any such change. While the observation that a moral norm does not always successfully guide action may prompt further thought about why this is so, it does not appear to provide any independent reason to doubt the viability of the norm in question.

Secondly, morality may fail to be action-guiding if it does not select a single unique action to be performed. There are a number of ways in which this may be the case. For example, a moral theory may rule out some options while leaving open multiple alternatives that are all deemed permissible. Alternatively, there may exist moral dilemmas in which a given moral view suggests either that there is no available option that is morally permissible, or that there are multiple options that are obligatory. Lastly, there may be cases that fall completely beyond the domain of a given moral theory, about which that theory is completely silent. In all of these cases, morality fails to pick out a single course of action to be pursued. The choice is not settled by moral considerations alone and the agent is left having to choose between multiple options, be they permissible, obligatory, prohibited or morally neutral. Whether or not these sorts of action-guidance failures are problematic for a view that exhibits them is a much-disputed topic in moral philosophy. I will not wade into this debate here, but want instead to point out that this is not the kind of action-guidance failure at stake in the topic at hand. Objectivism concerning moral beliefs does not entail that there is no uniquely right course of action to take, in any of the ways described above, since what an agent should do under moral uncertainty is simply what they should do if they had full moral knowledge and this may well pick out a single action. Of course, a given objective norm might not pick out a single action in any of the ways described, but that would be a problem for that particular norm, not for objectivism in general. One's view on the requirement for moral theories to select a unique course of

action to be pursued should not directly affect one's take on the debate between objectivism and subjectivism about moral beliefs.

Rather than either of these types of action-guidance failure, the difficulty for objectivism is that agents are unable to adhere to the view, at least some of the time, because they lack epistemic access to its prescripts. The problem lies not with the objective norms themselves, but with the broader objectivist moral view that incorporates them. This view cannot be applied under moral uncertainty if there are no further provisions (i.e. subjective norms) that guide action when an individual does not know which are the true objective norms. If what an agent ought to do does not depend on their moral beliefs, then there is no moral norm that is available to them when they are uncertain about some moral matter pertinent to their situation. This is because our moral beliefs are all we have. When making a choice under moral uncertainty, all Alex has to guide her are her partial moral beliefs, just like when Rosa is deciding whether or not to take her umbrella, all she has to guide her choice are her beliefs about the weather, and when Shawn is trying to decide what to do about his mother's medicine, all he has to guide his choice are his beliefs about the medicines. Therefore, morality itself cannot guide her beliefs; only her beliefs about morality can. This is related to a widely discussed objection to consequentialism: that it cannot guide our actions, since it is impossible to know in advance what consequences an action will have and, therefore, which action will have the best consequences. [35] What James Hudson said of utilitarianism holds true for objectivism: *"for human agents the theory is not really "action-guiding": it* does *say what one should do, but it gives this information in a practically unusable way."* [36]

[35] Hilary Greaves, "Cluelessness," *Proceedings of the Aristotelian Society* 116, no. 3 (2016): 311–39, https://doi.org/10.1093/arisoc/aow018; James Lenman, "Consequentialism and Cluelessness," *Philosophy & Public Affairs* 29, no. 4 (2000): 342–70, https://doi.org/10.1111/j.1088-4963.2000.00342.x; Alastair Norcross, "Consequentialism and the Unforeseeable Future," *Analysis* 50, no. 4 (1990): 253–56, https://doi.org/10.1093/analys/50.4.253; Tyler Cowen, "The Epistemic Problem Does Not Refute Consequentialism," *Utilitas* 18, no. 4 (2006): 383–99, https://doi.org/10.1017/S0953820806002172; Fred Feldman, "Actual Utility, The Objection from Impracticality, and the Move to Expected Utility," *Philosophical Studies* 129, no. 1 (2006): 49–79, https://doi.org/10.1007/s11098-005-3021-y; Dale Dorsey, "Consequentialism, Metaphysical Realism and the Argument from Cluelessness," *The Philosophical Quarterly* 62, no. 246 (2012): 48–70, https://doi.org/10.1111/j.1467-9213.2011.713.x; Joanna Burch-Brown, "Clues for Consequentialists," *Utilitas* 26, no. 1 (2014): 105–19.
[36] James Leonard Hudson, "Subjectivization in Ethics," *American Philosophical Quarterly* 26, no. 3 (1989): 221.

Now that we are clear about what is meant by a failure of action-guidance in the context of objectivism and subjectivism about moral beliefs, we are in a position to consider the extent to which this is really problematic after all. As we have seen, there are numerous ways in which a moral theory may fail to guide action, and a lack of action-guidance alone does not automatically undermine the plausibility of the theory under consideration. The question, therefore, is whether the case for objectivism with respect to moral beliefs is weakened by the particular way in which it fails to be action-guiding under moral uncertainty. The answer is that two potential issues seem to arise from this failure of action-guidance, but, upon further inspection, both can be overcome.

Firstly, we may find objectionable the way in which this view diminishes the role of morality in our lives. A lack of epistemic access to the correct moral norms means that they are unavailable to influence what one should do, but this is exactly the role that morality is supposed to play. It certainly sounds problematic if an account of morality leaves it no role in determining how morally conscientious agents act. However, it is not the case that failures of action-guidance leave *no* role for morality in our lives. Morality can be indirectly action-guiding under moral uncertainty, insofar as our beliefs directly guide our actions and these are aimed at moral truth. If, through careful consideration, our beliefs can track true morality, then morality can influence our actions via these beliefs, because rational agency involves acting on one's beliefs to achieve one's aims. In fact, this is a fairly simple idea that applies in all sorts of cases that involve no moral uncertainty at all. Recall Rosa's decision to take an umbrella. The weather does not directly guide her actions; rain does not enter her brain and cause her to choose the umbrella. The weather influences her beliefs, which, in conjunction with her desires, influence her choice. The true state of the world guides her choices only insofar as her beliefs are able to track the truth. Likewise, morality guides our actions only insofar as our moral beliefs track the truth and thereby inform our behaviour. Of course, sometimes this may lead agents to do the wrong thing, just as acting in accordance with one's empirical beliefs might lead one to take actions that do not give rise to the most desirable outcomes. But as long as we take our moral beliefs to be the best available guide to moral truth, then acting in accordance with them provides a crucial role for morality in influencing our action.

This coheres with accepted norms of epistemic rationality. There is some true set of facts about the world and, objectively, one should believe the truth. "*Believe the truth*" is not an action-guiding epistemic norm, since we are often uncertain about what is true. But this does not mean that truth plays no role in guiding our beliefs. There are subjective norms of epistemic rationality that dictate what we should believe given our limited evidence. These may occasionally lead to false belief if, for example, one has good evidence for $p$, but $p$ is actually false. But we take these norms to be the best guide to the truth given our limited evidence, nonetheless. Likewise in moral choices, subjective norms of rationality may sometimes suggest that one rationally ought to act in a way that is morally wrong, if one believes that some action is right when in fact it is not. But these norms are nonetheless the best guide to the right and the good, given our imperfect moral beliefs. Of course, for this analogy to hold, we need an account of moral epistemology that can explain why we should think that our moral beliefs are able to track moral truth. All I will say on this point for now is that the idea I want to emphasise—that morality can play a prominent role in our choices even if it is not directly action-guiding—is at least consistent with a range of different views on moral epistemology. And, of course, the sceptical challenge to the view that our beliefs can track the truth is not unique to moral beliefs.

The second potential problem for a view that suggests that morality is not action-guiding under moral uncertainty is that we find ourselves back to square one, facing the very problem we started with. We were seeking an answer to the question of what one to do when uncertain about what they ought to do. If there are instances in which there are no subjective moral norms to guide action under moral uncertainty, then we have no answer to this question. However, it should be noted that this is a practical problem for the morally uncertain agent, and the fact that a moral view leaves us facing a practical problem needn't always undermine the theoretical plausibility of that view. For example, suppose that morality prescribes donating a significant portion of one's income to charities that aim to ameliorate global poverty. Establishing this moral norm may leave an individual facing the practical problem of deciding how much money to donate, since they are somewhat torn between their moral commitments and their selfish desire for wealth. The presence of this practical problem for the individual does not call into question the moral norm itself. And remember that it is not objective moral norms per se that fail to be action-guiding, but the broader view of objectivism, just as it is not the prescription to donate that presents a

practical problem, but the broader motivational state of the agent in question. Of course, it might be ideal if moral theory was available to guide action in all situations, but it is not necessarily the case that individuals are always in possession of the information required in order to know what morality prescribes, and this may often leave them facing the practical challenge of how to act given their limited moral knowledge. One may claim that the very meaning of the term "morality" is an account of how best to act. Nonetheless, it is conceivable that there are objectively true moral norms to which we do not always have epistemic access, irrespective of whether or not one would continue to call this morality upon learning of its limitations.

Moreover, this practical problem is one that we already have a way of overcoming. Even if an objective account of morality is not able to guide our choices alone, the principles of instrumental rationality are on hand to fill the gap. Therefore, the practical challenge left by an account of morality that is not sensitive to agents' moral beliefs is neither a theoretical problem for the view itself, nor an insurmountable practical problem that stops rational agents from making choices under moral uncertainty in a coherent and morally conscientious manner.

## 6     Conclusion

Recall the dilemma presented at the end of §3: either moral ignorance always excuses wrongdoing, or morality fails to be action-guiding under moral uncertainty. I have shown how the divide between objectivism and subjectivism is in fact more complex and less worrisome than this dilemma suggests. Whether or not there are subjective moral norms, which dictate what an agent should do given their moral beliefs, depends on the details of the situation, such as the way in which the agent's moral beliefs were formed in the first place. In certain circumstances, we may endorse moral norms that are insufficiently sensitive to agents' moral beliefs to be fully action-guiding, but this needn't be as troublesome as it may at first appear. When morality fails to be action-guiding in this particular way, it can still play an indirect role in guiding our choices and the remaining question of deciding how to act poses a practical, rather than theoretical, challenge. In this

way, invoking the norms of instrumental rationality goes some way to resolving the theoretical dispute between objectivism and subjectivism under moral uncertainty.

Despite providing a guide to action without implying that moral ignorance always excuses wrongdoing, relying on norms of rationality under moral uncertainty has a potentially troubling consequence of its own: how one ought to act appears severed from what is morally right and wrong. If, as claimed above, morality is sometimes unable to guide action under moral uncertainty, it is possible that one could act according to the best available action-guiding norms, yet still do something that is morally wrong. That morality is not always action-guiding does not entail that it cannot provide an account of which actions are right and wrong. An act may be rational, given a particular set of moral beliefs, and yet be wrong if it contravenes the true moral norms.

This idea may not sit comfortably at first, as it seems natural to say that an agent does something wrong if and only if they do not act as they ought to act. However, this issue can be resolved by specifying more precisely what is meant by "ought" in this context. It may be true that an agent does something wrong if they do not act as they *morally* ought. Acting as one *rationally* ought, on the other hand, is no guarantee of moral rightness, unless one has complete knowledge of the true moral theory. Given that there are situations in which the only available action-guiding norms are those of rationality rather than morality, an agent may do something wrong even if they act as these norms dictate. Thus, the claim that an agent does something wrong if and only if they do not act as they ought may well be true for a moral interpretation of "ought", but not for rational interpretation of "ought". This coheres with the intuition that false moral beliefs are not always exculpatory. It may be rational for a gang member to inflict unwarranted harm on some individual, given their own particular moral beliefs and motivations, but this is immoral nevertheless. There are many acts that might be rational, for a particular agent, even if they are immoral. This is not to *endorse* such acts. We rightly criticise immoral conduct, irrespective of its instrumental rationality. Indeed, it is often the most calculated and efficient wrongdoers for whom we reserve the strongest moral criticism. The point is simply to distinguish that which an agent ought to do, in terms of the appropriateness of their means to their ends, from that which they ought to do, morally speaking.

One upshot of this is that, if false moral beliefs can make it instrumentally rational for morally motivated agents to act immorally, then they have good reason to think very seriously about moral questions. A morally conscientious agent should not simply be satisfied with acting rationally given their current moral beliefs, because of the possibility of being led astray if these beliefs are mistaken. To avoid acting immorally, they must aim for their moral beliefs to conform as closely as possible to the truth. This provides cause for reflecting on our moral judgements and studying moral philosophy, assuming of course that doing so makes one more likely to hold true moral beliefs.

The literature to date in the debate between objectivism and subjectivism under moral uncertainty suggests two possible options: either there are subjective norms under moral uncertainty, so acting in accordance with false moral beliefs excuses wrongdoing, or there are none, so morality is not action-guiding. As we have seen, however, moral ignorance is pertinent to the moral status of an agent's actions in some scenarios, and in others it is not, but this doesn't settle the question of whether there are *any* action-guiding norms that are sensitive to our moral beliefs. Even in situations in which one's moral beliefs do not determine what one morally ought to do, there are action-guiding norms of instrumental rationality that dictate how agents ought to go about achieving their particular ends, given their particular beliefs. Thus, we can see how individuals' actions may be guided by important moral considerations, without having to accept the view that false moral beliefs always excuse wrongdoing. This, of course, leaves open the question of what, precisely, these norms dictate. This is the question addressed in Chapter 2.

# CHAPTER 2

# Preference and Commitment Under Moral Uncertainty

## 1      Introduction

Recent years have seen something of a boom in philosophical work concerning the question of how agents should act under moral uncertainty. That is, what they should do when they do not know what they should do. A number of proposals for answering this question understand it as a challenge for rational choice. They are attempts to establish what agents *rationally* ought to do when they uncertain about what they *morally* ought to do. It is intriguing, therefore, that these accounts make little or no mention of agents' preferences or desires—notions that are central to traditional theories of rationality—but instead work entirely with the values provided by the moral theories about which the agent is uncertain. In so doing, these accounts fail to recognise the ways in which rational agents must strike a balance between their moral commitments and other, non-moral considerations, such as their own self-interest. In this paper I will argue that this feature is the root of a number of the most stubborn problems facing one prominent approach to rational choice under moral uncertainty. I will call this approach the Analogical View, since it treats moral uncertainty as analogous to empirical uncertainty. By finding a role for some conventional notion of preference or desirability, arrived at through a compromise between self-interest and moral commitment, I aim to modify this view in such a way as to avoid these problems.

This is a somewhat modest task: I do not aim to provide a full defence of the Analogical View, nor to compare it to all possible alternatives. Some readers will take issue with the view for reasons above and beyond those considered here. However, I at least aim to show that this modified approach fares better than existing instantiations of the Analogical View, by avoiding their most substantial flaws, while maintaining their major advantages.

§2 will characterise the Analogical View in more detail. §3 will explore three well-known objections to this approach: (i) swamping by 'fanatical' theories; (ii) the problem of

intertheoretic comparisons; and (iii) limited scope to numerically representable theories. §4 will suggest a unifying diagnosis of these theories and §5 will propose a remedy. §6 concludes the paper with some thoughts about the implications of this sort of formal decision procedure for choices made in the real world.

## 2    The Analogical View

Decision theory aims to describe, explain, appraise, and guide the choices of rational agents. Decision situations are typically characterised as involving a number of options, between which the agent must choose, and a number of possible states of the world, about which the agent is uncertain. Each particular combination of an option and a state of the world produces an outcome. Decision problems thought of this way can be conveniently represented by a matrix, with the available options as rows, the various possible states of the world as columns, and the outcomes that result from the combination of an option with a state of the world as elements. For example, if an agent is choosing whether to walk to a restaurant for a meal out, or stay at home for a less exciting dinner, but is uncertain about whether or not the restaurant will have a table available, the choice may be represented in the following matrix:

**Table 1.** Decision matrix for dinner choice

|  | Table available | No table available |
|---|---|---|
| Walk to the restaurant | Dinner at the restaurant and a walk | Dinner at home and a walk |
| Stay at home | Dinner at home and no walk | Dinner at home and no walk |

Which option is to be chosen depends on the agent's judgement of the probability of the various possible states of the world and the desirability, or value, of the different outcomes. These two judgements can be combined to give the expected value of each option. The value of each outcome is multiplied by the probability of the state of the world in which it would come about. The sum of these weighted values gives a probability-weighted average

of the possible values for a given option—its expected value. This generalises to any decision problem that can be represented as a combination of options $A_1$ to $A_m$, states of the world $S_1$ to $S_n$, and consequences $C_{11}$ to $C_{mn}$, in which an agent has a probability function $P$ over the states and a value function $v$ over the consequences. Assuming that the states are probabilistically independent of the options, such decisions can be represented by the following generic probability-value matrix:

**Table 2. P**robability-value matrix

|  | $P(S_1)$ | … | $P(S_n)$ |
|---|---|---|---|
| $A_1$ | $v(C_{11})$ | … | $v(C_{1n})$ |
| … | … | … | … |
| $A_m$ | $v(C_{m1})$ | … | $v(C_{mn})$ |

The expected value is given by:

$$EV(A_i) = \sum_{j=1}^{n} v(C_{ij}) \cdot P(S_j)$$

Expected utility theory (EUT) states that rational agents ought to choose the option with the highest expected value, relative to that agent's own judgement of the probability of the states and the desirability of the consequences. There is also a descriptive interpretation of EUT—that agents actually do maximise expected value—but the focus here shall remain on the normative claim. There are, of course, controversies and debates about the correct interpretation of this theory of rational choice, but something like EUT is generally accepted as the orthodox account of rational decision making under uncertainty about the state of the world. [37] However, this is not the only type of uncertainty facing decision-makers. Even if an agent knows exactly which outcome will result from each available option, they may still remain unsure of how to act, because they are morally uncertain. That is, they are uncertain about some moral matter that is relevant to their evaluation of the alternatives available to them.

---

[37] See §4 for more on the interpretation of EUT.

How agents ought to act under moral uncertainty has become a prominent point of contention in philosophy over recent years. Before discussing possible approaches to this problem, however, I wish to make three clarificatory points. First, "moral uncertainty" is often defined in contrast to "empirical uncertainty", but this may have some misleading implications. For one thing, it may suggest that moral and empirical are the only types of uncertainty. Empirical uncertainty is taken to mean the type of uncertainty about the state of the world discussed above: that is, degrees of belief divided between different ways that the world could be, which will determine the outcome of the option that is chosen. Agents may also be uncertain about which options are available to them, about which mutually exclusive and collectively exhaustive states of the world are pertinent to their choice, or about how to value the alternatives with respect to considerations other than morality. [38] Moreover, the moral-empirical distinction, along with the related distinction between normative and descriptive statements, seems to assume a stance on meta-ethical issues that I wish to remain neutral on for the time being. Proponents of naturalistic moral realism, for example, may consider normative claims to be descriptive of some moral reality, while certain views on moral intuitions hold that they provide empirical grounds for moral judgements. [39] Take, for example, the well-known claim made by W. D. Ross that, *"the moral convictions of thoughtful and well-educated people are the data of ethics just as sense-perceptions are the data of a natural science."* [40] For the sake of consistency with existing literature, I will stick to the standard terminology of moral versus empirical uncertainty, but note that this should not be taken to imply that these are the only types of uncertainty, nor that the discussion rules out any particular form of moral intuitionism. The distinction that matters is that between uncertainty about moral propositions and uncertainty about non-moral propositions.

Second, whatever terms we use, the distinction between moral and empirical uncertainty itself is not always so clear cut. Although uncertainty about the availability of a table at a restaurant seems straightforwardly empirical and uncertainty about, say, the action-omission distinction is clearly moral, there are many situations in which this distinction begins to break down. For example, uncertainty about the moral importance of animal

---

[38] Richard Bradley and Mareile Drechsler, "Types of Uncertainty," *Erkenntnis* 79, no. 6 (2014): 1225–48.
[39] Ross, *The Right and The Good*; Andrew Cullison, "Moral Perception," *European Journal of Philosophy* 18, no. 2 (2010): 159–75; Robert Audi, *Moral Perception* (Princeton: Princeton University Press, 2013).
[40] Ross, The Right and The Good, p.41

welfare might include moral considerations, such as the extent of our duty to prevent suffering, as well as empirical matters concerning the consciousness of non-human animals and their capacity for suffering. More generally, the idea of "*thick concepts*" blurs the boundary between the moral and the empirical. [41] For the sake of clarity, I will stick to cases that fall more clearly either side of the moral-empirical divide, but acknowledge that in real life this will rarely be so simple. How we ought to deal with complex mixtures of uncertainty remains an open research question for future work.

Lastly, it is important to clear up some ambiguity in the use of the word 'ought'. There are two meanings of the word that we must keep distinct when discussing what we ought to do under moral uncertainty. On the one hand, the term describes our first order moral obligations: what we morally ought to do in a range of different situations. One the other hand, the term also refers to a kind of second order norm that arises under uncertainty about first order moral theory. In other words, what we ought to do when we don't know what we morally ought to do. I take the meaning of 'ought' in this context to express norms of rationality, rather than morality. There are independent reasons for thinking that this is the right way to construe the second order norms, as I explored in Chapter 1, but even if you believe that there are some second order moral norms, we may still ask what rationality requires of us when we are uncertain about what we morally ought to do. That is the question at issue in this paper, to which we now turn.

Some have argued that we should treat moral uncertainty as roughly equivalent to empirical uncertainty, using the tools of decision theory. [42] Call this the Analogical View. The standard-bearer of the Analogical View is a procedure called Maximisation of Expected Choiceworthiness (MEC). [43] On this account, the states of the world in the decision matrix are replaced with the various moral theories in which an agent has some credence. Talk of 'theories' is perhaps misleading; it is really just shorthand for any moral consideration in which one holds some credence, not only comprehensive moral theories. The agent's moral beliefs take the form of a probability distribution over the different moral views

---

[41] Bernard Williams, *Ethics and the Limits of Philosophy* (Abingdon: Routledge, 1985).

[42] See, for example: Lockhart, *Moral Uncertainty and Its Consequences*; Ross, "Rejecting Ethical Deflationism"; Andrew Sepielli, "What To Do When You Don't Know What To Do," in *Oxford Studies in Metaethics, Volume IV*, ed. Russ Shafer-Landau, vol. 4 (Oxford: Oxford University Press, 2009), 5–28; MacAskill and Ord, "Why Maximize Expected Choice-Worthiness?"; Christian Tarsney, "Intertheoretic Value Comparison: A Modest Proposal," *Journal of Moral Philosophy* 15, no. 3 (2018): 324–44.

[43] MacAskill and Ord, "Why Maximize Expected Choice-Worthiness?"; MacAskill, Bykvist, and Ord, *Moral Uncertainty*.

under consideration. Each theory provides an evaluation of each option, which is referred to as its choiceworthiness and is represented by a numerical value. These choiceworthiness values replace the agent's judgement of the desirability of the outcomes in the conventional probability-value matrix. Choiceworthiness is the all-things-considered evaluation of an option according to a given moral view and need not only concern the consequence of the choice. This is to allow MEC to handle moral theories that do not take consequences to be the sole or principal locus of moral evaluation. An option's choiceworthiness may be evaluated in light of the consequences it gives rise to, features of the option itself, or the qualities (virtues and vices, perhaps) of the agent that are expressed by choosing it. Thus, this approach is not committed to any particular claim about what type of moral theory is correct and can accommodate a wide range of views. So, a choice under moral uncertainty involves options $A_1$ to $A_m$, moral theories $T_1$ to $T_n$, a probability function $P$ over the moral theories, and a set of choiceworthiness functions $cT_1$ to $cT_n$, which assign choiceworthiness values to the options according to the theory under consideration. This can be represented by a probability-value matrix (Table 3) and the expected choiceworthiness of each option can be calculated in much the same way that a choice is represented and expected value is calculated according to orthodox decision theory under empirical uncertainty.

**Table 3.** Probability-value matrix under moral uncertainty

|          | $P(T_1)$   | …   | $P(T_n)$   |
|----------|------------|-----|------------|
| $A_1$    | $cT_1(A_1)$ | …   | $cT_n(A_1)$ |
| …        | …          | …   | …          |
| $A_m$    | $cT_1(A_m)$ | …   | $cT_n(A_m)$ |

$$EC(A_i) = \sum_{j=1}^{n} cT_j(A_i) \cdot P(T_j)$$

The claim made by proponents of MEC is that, insofar as choices that maximise expected value are rational under empirical uncertainty, choices that maximise expected choiceworthiness are rational under moral uncertainty. A simplified example can demonstrate how this is supposed to work and why we might find it appealing. Consider the numerous examples of lottery winners who have wanted to 'do the right thing' and

give some of their winnings away. Suppose that just such a person had decided on a fixed sum that they wanted to give away, but was uncertain of who would be the most morally worthy recipient. On the one hand they think that it might be best to give the money to whichever charity could be shown most effectively to promote wellbeing and alleviate suffering. On the other hand, they give some credence to the view that they have special obligations to help their close family. Suppose that according to the first view, call it $T_1$, it would be much better to donate the money to Malaria Consortium, since it is regarded as the most effective charity around. [44] This view could be represented with the choiceworthiness function $cT_1$ such that $cT_1$(Malaria Consortium)=1000 and $cT_1$(Family)=10. However, according to the second view, call it $T_2$, the preferential weighting of benefits given to one's nearest and dearest means that it would be slightly better to use the money to provide financial security for a number of close relatives. This view could be represented with the choiceworthiness function $cT_2$ such that $cT_2$(Malaria Consortium)=40 and $cT_2$(Family)=60. This decision between donating to Malaria Consortium and helping one's family under uncertainty about $T_1$ and $T_2$ may be characterised as follows:

**Table 4.** Decision matrix for charitable giving

|  | $T_1$ | $T_2$ |
|---|---|---|
| Give to Malaria Consortium | 1000 | 40 |
| Give to family | 10 | 60 |

Suppose that our lottery winner thinks that the family-oriented moral view $T_2$ is somewhat more likely to be correct than the impartial, utilitarian alternative $T_1$, such that their credence is split roughly 60:40. The expected choiceworthiness of these options would then by given by:

$$EC(\text{Malaria Consortium}) = (1000 \cdot 0.4) + (40 \cdot 0.6) = 424$$

$$EC(\text{family}) = (10 \cdot 0.4) + (60 \cdot 0.6) = 40$$

---

[44] GiveWell, "Our Top Charities," November 2020, https://www.givewell.org/charities/top-charities.

According to MEC, therefore, they should donate the money to Malaria Consortium, since it has the higher expected choiceworthiness.

This approach has some appealing features. For one thing its structural similarity to standard EUT gives a parsimonious and consistent account of rational choice under uncertainty. If EUT is the correct approach to rational choice under empirical uncertainty, then it seems as though something similar should be considered rational under moral uncertainty. There are many different types of proposition about which we may be uncertain and we do not adopt different decision procedures for each. We may be uncertain about the weather, about upcoming election results, or about the future performance of financial markets, but we do not adopt different decision theories for meteorological, political, or economic uncertainty. Absent further argument, we should not treat moral uncertainty as substantially different from uncertainty about any other kind of proposition. [45]

Another, perhaps more substantial, advantage of maximising expected choiceworthiness is that it is sensitive to how much is stake according to different moral theories. The degree of influence that a theory has over a choice should be proportional to both the likelihood of that theory being correct and the amount that is judged to be at stake in the choice according to that theory. This means that if you are fairly confident that option $A$ is slightly better than option $B$, but give some credence to a view according to which $A$ is much worse, it seems as though you ought to give this low credence, high stakes view some sway. For example, if I believe that eating meat is probably morally acceptable, but that there is a small chance that it is severely morally wrong, then it would seem sensible to avoid taking the considerable moral risk of eating meat for the comparatively small benefit of a slightly tastier meal. Analogously, even if you think it is much more likely that you won't be involved in a car crash than that you will, you still wear a seatbelt, because if that did happen, the stakes would be far higher. This choice would be rational and is exactly what EUT would prescribe, given plausible value and probability assignments. Sensitivity to stakes is a result of the fact that both probability and value factor into the calculation of expected choiceworthiness. Therefore, even if one moral view is less likely to be correct than another, it may contribute more to the difference in the expected choiceworthiness

---

[45] MacAskill and Ord, "Why Maximize Expected Choice-Worthiness?"

of the options if the difference in the value of the outcomes according to that theory are sufficiently large. In other words, if one theory suggests that more is at stake in a choice than another theory then, *ceteris paribus*, it should have a greater influence over our choice.

This is a feature that other approaches to choice under moral uncertainty lack. For example, some have argued that it is only the true moral theory, rather than an agent's beliefs about morality, that determine how they should act. [46] Thus, moral uncertainty is deemed to be irrelevant to what such an agent ought to do. According to this view (known variously as objectivism, actualism, and normative externalism) the fact that the agent in question gives some credence to a view according to which *A* is much better than *B* is entirely irrelevant, if *B* is in fact better than *A*. This approach is neither sensitive to stakes nor to degrees of belief, since moral beliefs play no role in determining how individuals ought to act. Furthermore, it fails to be of any use to an agent under moral uncertainty: they cannot follow the rule simply to 'do the right thing', since this is precisely what they don't know. It is like the demand to believe the truth or to pick the wining lottery numbers.[47]

Some approaches do accommodate a role for moral belief in choice under moral uncertainty, but nonetheless fail to exhibit sensitivity to stakes. One such approach suggests that agents ought to do that which they believe is most likely to be right. [48] In other words, they should settle on the moral theory that they deem most likely to be correct and then act in line with whatever it recommends. This approach, known as My Favourite Theory (MFT), might seem to be a reasonably realistic characterisation of the way that many people make moral decisions. It factors in agents' moral beliefs, since what they ought to do is determined by the theory in which they have the greatest degree of belief. However, it is insensitive to stakes, since it does not consider the content of any theories other than that which is deemed most likely. In the example considered above, MFT would recommend eating meat, because this is the best option according to the moral view with the highest subjective probability. The fact that the choice is far higher stakes according to a second favourite theory that recommends a different course of action has no bearing on what this person ought to do according to MFT. In this case, it seems wrong to completely ignore the slightly less likely moral view, because if this view were true then one of the

---

[46] Weatherson, "Running Risks Morally"; Harman, "The Irrelevance of Moral Uncertainty."
[47] See Chapter 1 for a more detailed discussion of this view.
[48] Gustafsson and Torpman, "In Defence of My Favourite Theory."

options would involve serious moral wrongdoing. But, according to MFT, this is exactly what we should do. Thus, sensitivity to stakes is a substantial advantage of Analogical Views, such as MEC, over stakes-insensitive views, such as actualism or MFT.

Despite these appealing features of the Analogical View, there are number of well-known and substantial objections to this approach. It is these to which we now turn our attention.

# 3　　Three Problems for the Analogical View

I will consider three main problems for the Analogical View. The first two are closely related to the fact that calculating the expected choiceworthiness of an option requires value comparisons across the different moral theories under consideration. More specifically, we must be able to compare the magnitude of value differences between options according to one theory with the magnitude of value differences between options according to another. For example, consider a decision between option $A$ and option $B$, under uncertainty about moral theories $T_1$ and $T_2$. Suppose that $T_1$ ascribes a higher value to $A$ than $B$, but $T_2$ ascribes a higher value to $B$ than $A$ and the agent in question has a credence of 0.5 in each view. Given that the agent gives equal credence to these two theories, if the difference between $A$ and $B$ according to $T_1$ is equal to the difference between $B$ and $A$ according to $T_2$, then the agent should be indifferent between the options. In order for $A$ to have higher expected choiceworthiness than $B$, the value difference between $A$ and $B$ according to $T_1$ must be greater than the value difference between $B$ and $A$ according to $T_2$. Furthermore, if $T_2$ was judged to be twice as likely to be correct as $T_1$, then the value difference between $A$ and $B$ according to $T_1$ would have to be greater than double the value difference between $B$ and $A$ according to $T_2$ in order for $A$ to come out with the higher expected choiceworthiness. These intertheoretic value comparisons are essential in order for the machinery of EUT to get up and running in choices under moral uncertainty.

## 3.1　　Limited Scope

The first problem that arises from the need for intertheoretic value comparisons is that the scope of the Analogical View appears to be severely limited to only those choices in which

all of the moral theories under consideration can be numerically represented in a particular way. Calculation of expected choiceworthiness is only possible if the theories under consideration can be thought of as ascribing to options values that are measurable on an interval scale. An interval scale allows ratios of value differences to be expressed. For example, the difference in value between option *A* and option *B* is double the difference between option *C* and option *D*. Although this structure is not required in order for a single moral theory to guide action in the absence of moral uncertainty, the use of expected choiceworthiness under moral uncertainty requires an interval scale in order to make the sort of value difference comparisons described above. Some moral theories, such as certain forms of utilitarianism, will provide the morally uncertain agent with nice, well-behaved numbers with which to work. However, many prominent alternatives, including classical theories of deontological, contractualist and virtue ethics, provide no such values.

This needn't be a devastating blow (yet), since a moral theory that does not directly ascribe values to options in this way may nonetheless be represented as though it does. These theories may provide an ordinal ranking of the options and this ordering may be represented by values that are measurable on an interval scale as long as it ranks risky prospects and satisfies the von Neumann-Morgenstern axioms of Transitivity, Completeness, Continuity, and Independence. [49] As long as all theories in question are capable of providing such an ordering, then we can calculate expected choiceworthiness. However, many moral views do not even provide this sort of ordering. For example, the Ten Commandments are a set of prescriptions and proscriptions, with no in-built value ascriptions or even ordering of options. These moral laws do not say that remembering the Sabbath has a value of 10 and coveting they neighbours wife has a disvalue of $-50$. Nor do they imply any particular ordering of options in terms of the degree to which the Commandments are satisfied. Therefore, any agent who holds some degree of belief in the Ten Commandments will be unable to calculate expected choiceworthiness for the options from which they must choose. And the Ten Commandments are not alone here; numerous moral theories are simply not in the business of either ordering or valuing options. Instead, they may concern only the moral character of the agent, the deontic status of their actions or the ways in which they relate to other people. Therefore, the requirement that all

---

[49] John von Neumann and Oskar Morgenstern, *Theory of Games and Economic Behavior* (Princeton: Princeton University Press, 1944).

theories under consideration are representable by values measurable on an interval scale severely limits the scope of the Analogical View.

## 3.2    The Problem of Intertheoretic Comparisons

The second problem is that even when all moral theories under consideration are representable on an interval scale, intertheoretic comparisons can still not meaningfully be made. This is known as the Problem of Intertheoretic Comparisons. There are two reasons for thinking that such intertheoretic comparisons are not possible, which give rise to two different versions of the problem. The first is that some different moral theories refer to fundamentally different conceptions of moral value and it does not seem that comparisons between such values can meaningfully be made. The issue is not that these comparisons are possible, but we don't know the exchange rate between, say, rights and hedons. Rather, there is no exchange rate to be known. None of these moral theories contains any information regarding its conception of value in the terms of the others. Nor is there some third theory, general scale or common measure that can be used to convert the units of one theory into the units of another. Such comparisons are simply meaningless, like trying to compare the aesthetic value of Picasso's Guernica with the epistemic value of knowing that it is displayed at the Museo Reina Sofía, or the height of a tree with the greenness of its leaves. Rights and hedons, just like aesthetic and epistemic value, or height and colour, are different entities and their quantities cannot meaningfully be compared. The tools of EUT simply cannot process these immiscible conceptions of moral value within the same decision. Call this the Reference Problem.

In fact, this problem renders intertheoretic value comparisons even more deeply problematic than comparisons of other different kinds of quantity. Given that the rival moral theories are mutually inconsistent, only one of them can be true. Therefore, only one of the competing conceptions of moral value can truly exist, whether consisting of absolute moral laws, virtues, units of welfare, or something else. Any comparison between different conceptions of moral value must be false, since at least one of the entities being compared must not exist. Even if one knew everything there was to know about each theory in question, one would be no closer to being able to compare quantities of different conceptions of moral value. This would be like trying to determine the exchange rate between Japanese yen and Spacebucks, the fictional currency in the 1987 sci-fi cult classic,

*Spaceballs.* This is not just a matter of finding out the true exchange rate, like that between yen and Kenyan Shillings. Rather, no exchange rate between yen and Spacebucks can ever be found, because Spacebucks themselves do not really exist. We may not know for sure which conception of moral value truly exists, but given that not all of them do, there is simply no fact of the matter about how quantities of these values compare.

The Reference Problem applies when attempting to make comparisons of value across moral theories that substantially differ in their explanation of the nature of moral value. However, sometimes we may only be uncertain about a fairly small detail of moral theory and are attempting to make value comparisons between theories that agree on the fundamental nature of moral value. For example, one may be certain that some form of prioritarianism is the correct account of how to distribute scarce healthcare resources, but remain uncertain about the precise weighting of benefits given to people at different levels of welfare. Alternatively, one may be fully convinced by the arguments for critical level theories of population axiology, but be uncertain of where to set the critical level. In these sorts of cases, it seems as though the Problem of Intertheoretic Comparisons is not a problem at all.

However, there is a second reason for thinking that intertheoretic comparisons are not possible. As mentioned above, we require values that are measurable on an interval scale in order to make the comparisons between value differences for calculating expected choiceworthiness. Some moral theories may provide values with this structure, but interval scales bring with them a problem of their own. An interval scale is uniquely determined only up to positive affine transformations. If $T$ is a function producing an interval scale, then $T^*$ is a positive affine transformation of it if and only if it takes the form $T^* = aT + b$, where $a$ is a positive constant and $b$ is any constant. The problem for intertheoretic value comparisons is that the values assigned by each theory under consideration may be altered by different positive affine transformations. Different transformations will lead to different outcomes when attempting to calculate expected choiceworthiness and there is no way of establishing a 'true' or 'correct' scale for each moral theory. Given that each theory can only provide information about the ordering of outcomes and the ratios of value differences between outcomes, we have no way of calibrating the scale of each moral theory. Call this the Scale Problem.

For example, in the lottery winner case described above, I suggested choiceworthiness functions $cT_1$ and $cT_2$, but we could just as well represent these theories with choiceworthiness functions $cT_1* = 0.1(cT_1)$ and $cT_2* = 100(cT_2)$, which are positive affine transformations of the originals. These functions give the values in Table 5:

**Table 5.** Decision matrix for charitable giving with transformed values

|  | $T_1$ | $T_2$ |
|---|---|---|
| Give to Malaria Consortium | 100 | 4000 |
| Give to family | 1 | 6000 |

The expected choiceworthiness of these options would then be given by:

$$EC(\text{Malaria Consortium}) = (100 \cdot 0.4) + (4000 \cdot 0.6) = 2440$$
$$EC(\text{family}) = (1 \cdot 0.4) + (6000 \cdot 0.6) = 3600.4$$

Now giving to one's family has the higher expected choiceworthiness so, according to MEC, this is what should be done. This has reversed the recommendation of what to do as a result of switching between representations of moral theories that should be informationally equivalent. There is nothing within the theories themselves that can tell us which of these different choiceworthiness functions to use. Therefore, MEC can recommend one option or the other, depending on which choiceworthiness functions we use to represent the theories. And that, ultimately, is no recommendation at all.

The Scale Problem applies even when one is only uncertain about the finer details of a moral theory, having settled conclusively on the overall class of moral theory that must be correct. Even the closely related theories of prioritarianism with different weightings or population axiologies with different critical levels might undergo different transformations, so intertheoretic comparisons are not possible. Note, however, that this version of the problem of intertheoretic comparisons does not deny the fundamental meaningfulness of intertheoretic value comparisons, but is a practical problem for implementing MEC. Intertheoretic value comparisons may be fundamentally possible, but we just lack the

knowledge of how to correctly fix the scales of different theories. We may, therefore, try to find ways of fixing the scales, in order to avoid this problem. [50]

## 3.3    Swamping by Fanatical Theories

The third problem for the Analogical View is that calculations of expected choiceworthiness can be swamped by 'fanatical' moral theories. The trouble here is that, if theories that posit larger values have more influence over a choice, then the decision procedure can be dictated by some moral theory that seems highly implausible, but ascribes wildly large differences in value between the options. This problem and sensitivity to stakes are in fact two sides of the same coin: because expected choiceworthiness is sensitive to stakes, a highly dubious theory can swamp all other (more likely) theories under consideration, so long as it posits large enough values. [51] It seems problematic to think that a moral view that we are highly confident is false should determine what we ought to do, simply because it suggests huge value differences between the options. But under MEC, this is exactly what we get.

This is a version of a more general problem for EUT that has become known as Pascal's Mugging. [52] This is a thought experiment in which a mugger approaches Pascal in a dark alley and demands his wallet, containing 10 livres, in exchange for 20 livres the next day. Pascal refuses and the mugger responds by iteratively increasing the value of the promised repayment until he is offering to use his magic powers to grant Pascal an additional 1,000 quadrillion days of happiness, in exchange for taking the wallet now. He arrives at this value, having established that Pascal gives a probability of 1 in 10 quadrillion to the possibility that the mugger's promise will be kept. Since the expected outcome of handing over the wallet, 100 additional days of happiness, is worth more to Pascal than the 10 livres,

---

[50] For examples of attempts to do this see: Ross, "Rejecting Ethical Deflationism"; Lockhart, *Moral Uncertainty and Its Consequences*; Andrew Sepielli, "'Along an Imperfectly-Lighted Path': Practical Rationality and Normative Uncertainty" (Doctoral Thesis, Rutgers University, 2010); William MacAskill, "Normative Uncertainty" (DPhil thesis, University of Oxford, 2014).

[51] Brian Hedden, "Does MITE Make Right? On Decision-Making under Normative Uncertainty," in *Oxford Studies in Metaethics Volume 11*, ed. Russ Schafer-Landau (Oxford: Oxford University Press, 2016), 102–28; MacAskill, Bykvist, and Ord, *Moral Uncertainty*; Krister Bykvist, "Moral Uncertainty," *Philosophy Compass* 12, no. 3 (2017); MacAskill and Ord, "Why Maximize Expected Choice-Worthiness?"; MacAskill, "Normative Uncertainty."

[52] Eliezer Yudkowsky, "Pascal's Mugging: Tiny Probabilities of Vast Utilities," LessWrong.com, 2007, https://www.lesswrong.com/posts/a5JAiTdytou3Jg749/pascal-s-mugging-tiny-probabilities-of-vast-utilities; Nick Bostrom, "Pascal's Mugging," *Analysis* 69, no. 3 (2009): 443–45, https://doi.org/10.1093/analys/anp062.

he accepts the mugger's proposal and gives him the wallet. Of course, the unwritten epilogue is that the mugger does not keep his promise, nor did he ever intend to, and makes off happily with Pascal's wallet. The idea is that as long as one of the possible outcomes of an option is sufficiently valuable, then EUT will recommend taking it, even if it is almost certainly not going to be realised. But of course, this is not a sensible decision for Pascal to make and any theory of rational choice that prescribes it must have gone astray somewhere. Likewise, if MEC can lead agents to be 'mugged' in this fashion, by highly unlikely theories that posit vast values, it must not be the correct account of rational choice under moral uncertainty.

While insensitivity to stakes was considered a relative disadvantage of actualism and MFT, it becomes a relative advantage over the Analogical View in light of the problem posed by fanatical theories. According to actualism, one should never do what is required by an outlandish and unlikely moral theory, unless of course that theory happens to be true. Whatever its limitations regarding action-guidance, actualism at least has the advantage of never recommending that an agent should do something that would in fact be morally wrong. MFT, on the other hand, would recommend an immoral action if the theory that one judged most likely to be true was in fact false. Nonetheless, it is not beholden to the judgements of highly improbable moral theories that posit vast value differences, since these would, by definition, not be the 'favourite' theory of the agent in question. The same cannot be said of MEC.

## 4      A Unifying Diagnosis

There is a curious feature of MEC as an instantiation of the Analogical View: it is presented as a theory of rational choice, but it makes little or no mention of agents' ends, desires, or preferences—notions central to conventional accounts of rationality. There are two reasons for thinking that an agent's own preferences or desires should play an important role when discussing theories of rational choice. For one thing, these theories are traditionally considered to be accounts of instrumental rationality. [53] That is, claims about what would be the best way for an individual to go about achieving their aims or desires,

---

[53] For a detailed critique of this view, see: Johanna Thoma, "Advice for the Steady: Decision Theory and the Requirements of Instrumental Rationality" (PhD Thesis, University of Toronto, 2017).

whatever those may be. [54] A theory of morality may be able to ignore the particular ends of a person and still tell them what they ought to do; it doesn't matter whether you want to tell the truth or not, doing so is morally required. Theories of rationality, however, depend crucially on our own desires and beliefs. *If* you are hungry and believe that there is food in the kitchen, *then* it is rational for you to get up and walk there. If you have no such desire, however, then a theory of rationality offers no such prescription. This means-end conception of rationality extends at least as far back as Hume's famous adage that "*reason is, and ought only to be the slave of the passions.*" [55]

Secondly, preferences play a central role in the very account that MEC is an attempt to emulate: expected utility theory. Different interpretations of utility have given rise to two different ways of understanding the demands of EUT: mentalism and behaviourism. [56] According to a mentalist interpretation, utility is a real, psychological phenomenon, which may or may not be accessible by introspection. Early proponents of this view, such as Bernoulli, [57] Bentham, [58] and Mill, [59] conceived of utility as an agent's happiness and the absence of suffering. More recent accounts present utility as a measure of something like our folk psychological notion of desire. [60] Given mentalism about utility, the prescription to maximise expected utility amounts to the claim that rational agents ought to make choices that maximise the expected quantity of this psychological measure. This assumes that the relevant aspect of our psychology is accessible by introspection, is quantifiable, and can be measured on an interval scale.

---

[54] James Joyce, *The Foundations of Causal Decision Theory* (Cambridge: Cambridge University Press, 1999); Lara Buchak, "Risk and Tradeoffs," *Erkenntnis* 79, no. S6 (2014): 1091–1117.

[55] (Hume, 1739, p.415)

[56] José Luis Bermúdez, *Decision Theory and Rationality* (Oxford: Oxford University Press, 2009), https://doi.org/10.1093/acprof:oso/9780199548026.001.0001/acprof-9780199548026-chapter-3?print; Philip Pettit, "Decision Theory and Folk Psychology," in *Essays in the Foundations of Decision Theory*, ed. Michael Bacharach and Susan Hurley (Oxford: Blackwell, 1991), 147–75; Lara Buchak, *Risk and Rationality* (Oxford: Oxford University Press, 2013); Franz Dietrich and Christian List, "Mentalism Versus Behaviourism in Economics: A Philosophy-of-Science Perspective," *Economics and Philosophy* 32, no. 2 (2016): 249–81; Samir Okasha, "On the Interpretation of Decision Theory," *Economics and Philosophy* 32, no. 3 (2016): 409–33; Johanna Thoma, "Decision Theory," in *The Open Handbook of Formal Epistemology*, ed. Richard Pettigrew and Jonathan Weisberg (PhilPapers Foundation, 2019), 57–106.

[57] Daniel Bernoulli, "Exposition of a New Theory on the Measurement of Risk (1738)," *Econometrica* 22 (1954): 23–26, https://doi.org/10.1017/S0266267115000346.

[58] Jeremy Bentham, *An Introduction to the Principles of Morals and Legislation (1789)* (New York: Dover Publications, 2007).

[59] John Stewart Mill, *Utilitarianism (1861)*, ed. Roger Crisp (Oxford: Oxford Univserity Press, 1998).

[60] Richard Jeffrey, *The Logic of Decision*, 3rd ed. (Chicago: University of Chicago Press, 1965); David Lewis, "Radical Interpretation," *Synthese* 27, no. 3 (1974): 331–44; Pettit, "Decision Theory and Folk Psychology."

Behaviourism is the view, especially common among economists, that utility is not a real, psychological quantity, but is just a convenient mathematical construct that can be used to represent agents' preferences if they satisfy the axioms of a given representation theorem. These representation theorems aim to show that, for any agent whose preferences satisfy a certain set of axioms, there exists a utility function and a probability function such that the agent's choices can be represented as maximising expected utility. [61] According to this interpretation of utility, the norm of maximising expected utility is better understood as the claim that a rational agent ought to have coherent preferences, such that they can be represented as an expected utility maximiser, where coherence is defined by axioms of the representation theorem. In other words, it provides rational constraints on preferences, such as Transitivity, Completeness, and Independence, rather than making any claim about the most appropriate way of satisfying one's desires. The behaviourist conception of EUT is more permissive than the mentalist conception, since any set of choices that consistently maximise the expectation of some psychological measure will be representable as expected utility maximisation, but some choices that could be so modelled will not maximise satisfaction of one's actual desires. [62] For example, even if an agent made choices that *minimised* expected happiness, they could be represented as maximising expected utility on the behaviourist interpretation.

There is significant disagreement among philosophers over the correct interpretation of utility and, hence, EUT. No matter which interpretation we adopt, however, one thing remains the same: the centrality of agents' own preferences to theories of rational choice. In contrast to these conventional accounts, MEC works with choiceworthiness values that represent the evaluation of alternatives according to different moral theories. Its proponents briefly make the connection between moral choiceworthiness and individual preference with the notion of moral conscientiousness. [63] The idea seems to be that a morally conscientious agent will have a utility function that simply tracks the choiceworthiness function of whichever moral theory is being considered. But the relation between morality on the one hand and desire, preference, or utility, on the other is far

---

[61] von Neumann and Morgenstern, *Theory of Games and Economic Behavior*; Leonard Savage, *The Foundations of Statistics* (New York: Dover, 1954); Jeffrey, *The Logic of Decision*; Ethan Bolker, "Functions Resembling Quotients of Measures," *Transactions of the American Mathematical Society* 124 (1966): 292–312; Ethan Bolker, "A Simultaneous Axiomatisation of Utility and Subjective Probability," *Philosophy of Science* 34 (1967): 333–40.
[62] Okasha, "On the Interpretation of Decision Theory."
[63] MacAskill, Bykvist, and Ord, *Moral Uncertainty*.

more complex than this conception of moral conscientiousness allows. There is a rich history of philosophical debate about this relation and it has important implications for rational choice under moral uncertainty, as I aim to show in what follows.

In his 1977 paper *Rational Fools*, Amartya Sen criticised conventional EUT for its perceived inability to accommodate behaviour motivated by considerations other than agents' self-interest. He differentiated between several kinds of consideration that might influence a choice: narrow self-interest, concerning the way in which one's welfare depends only on the goods one consumes; sympathy, concerning the way in which one's welfare is affected by the welfare of others (including antipathy, if one's welfare is positively affected by the suffering of others); and commitment, when one's behaviour is motivated by something other than one's own welfare. Sen neatly captures the distinction with the following example:

> *"If the knowledge of torture of others makes you sick, it is a case of sympathy; if it does not make you feel personally worse off, but you think it is wrong and you are ready to do something to stop it, it is a case of commitment."* [64]

By distinguishing between narrow self-interest, sympathy and commitment, Sen is highlighting two points: first, our welfare may be determined by something other than our own consumption, such as when one is made to feel bad by witnessing the suffering of others; and second, our choices may be motivated by something other than our own welfare, such as a commitment to moral principles. Note that I am focussing specifically on moral commitments here, but Sen considers a broader range, including religious and political commitments, commitment to the interests of a particular group, or, relatedly, commitment as a solution to collective action problems like voting and the Prisoners' Dilemma.

Sen says that sympathy can be straightforwardly captured by EUT, since one's preference ordering reflects all factors that are expected to affect one's welfare, including the impact of others' welfare on one's own. Committed choices, however, are taken to be instances of counter-preferential choice. He claims that each agent has their own, true preferences,

---

which reflect their welfare under different alternatives, and that commitments motivate us to adopt different preference orderings in practice. This alternative preference ordering is arrived at through some process of compromise between one's preferences and one's moral commitments (described elsewhere as a compromise between one's own goals and the goals of others). [65] These counter-preferential choices are not accommodated by EUT, which mandates and predicts choice in line with agents' preferences. Sen says that commitment *"drives a wedge between personal choice and personal welfare, and much of traditional economic theory relies on the identity of the two."*

Insofar as committed choices are observed and deemed rationally permissible, they undermine both the descriptive and normative claims of conventional EUT. Individuals' commitments may be stronger or weaker, resulting in choices that deviate to a greater or lesser extent from their own preferences and this should not be considered irrational.

Daniel Hausman has defended rational choice theory from Sen's critique by arguing for a broader conception of preferences. [66] He claims that an agent's preferences should be thought of as all-things-considered evaluative judgements, which incorporate narrow self-interest, sympathy, commitments, and anything else that is deemed relevant by the lights of the agent in question. He agrees with Sen's claim that models of rational choice should be sensitive to a broader range of considerations than merely narrow self-interest, but thinks that this can be achieved by a single, richer conception of preference. In other words, the preferences that you end up adopting through compromise between your own welfare-regarding preferences and your commitments simply *are* your preferences. [67]

Although it seems natural to say, in accord with Hausman, that the preference ordering upon which an agent acts is in fact that agent's preference ordering, I will not attempt to defend that view here. Instead, I want to highlight the points of agreement between the opposing parties on this matter: that agents undertake some process of compromise between their own self-interest and their moral commitments when comparing alternatives to arrive at the preference ordering on which they will act. For present purposes, nothing

---

[65] Amartya Sen, *Rationality and Freedom* (Cambridge: Hardvard University Press, 2002); Philip Pettit, "Construing Sen on Commitment," *Economics and Philosophy* 21, no. 1 (2005): 15–32, https://doi.org/10.1017/S0266267104000367.
[66] Daniel Hausman, "Sympathy, Commitment, and Preference," *Economics and Philosophy* 21, no. 1 (2005): 33–50.
[67] For an alternative line of criticism, see also: Pettit, "Construing Sen on Commitment."

much hinges on whether we consider these preferences the agent's own, or think of this as a kind of counter-preferential choice.

Note also that a great deal is being left unsaid about the compromise between self-interest and moral commitment. This process is the subject of a number of contentious points in the theorisation of moral motivation, which lie beyond the purview of this paper. Let it suffice to say that, absent further argument, we should not assume that agents are rationally required to have any particular degree of commitment, nor therefore to arrive at any particular compromise between considerations of self-interest and morality, when making morally relevant choices. If I am less moved by certain moral considerations than another person, so end up adopting overall preferences that are more closely aligned to my narrowly self-interested preferences, it does not seem as though I am being less rational. You might criticise my level of commitment from a moral standpoint, but this is a separate matter.

With these ideas in mind, let us turn our attention back to the Analogical View and its prime exemplar, MEC. It seems on the face of it as though its proponents have in mind an agent for whom moral commitments entirely determine the preference ordering that is adopted, with no ground ceded in a compromise between morality and self-interest. If an agent was certain of a particular moral theory, MEC says that they would be irrational unless they acted exactly as this theory prescribed. Recall that the proponents of MEC claim that it is the right account of rational choice for morally conscientious agents, where this is taken to mean that they *"prefer doing right to doing wrong and are indifferent between different right-doings"*. [68] Now, as I have suggested, this sort of moral motivation should not be taken to be a rational requirement without further argument. And nor is it a very realistic characterisation of living, breathing people. Moral conscientiousness is not a binary notion, but rather comes in degrees. Moral conscientiousness may be thought of as the degree of compromise an agent is willing to make between self-interest and their moral commitments, in so far as these diverge. The degree of moral conscientiousness between individuals is highly variable and we should neither assume, nor require that agents' motivations are solely moral. So, either MEC qua theory of rationality is too demanding, by *requiring* agents to care only about morality, or it is unrealistic, by *assuming* that agents only care about morality.

---

[68] MacAskill, Bykvist, and Ord, *Moral Uncertainty*, p.20.

Perhaps this is simply an idealising assumption in order to isolate the question of how to make choices under moral uncertainty from other complicating factors. However, even if we think only of agents who are completely morally conscientious, there remains another crucial difference between MEC and the approaches of Sen and Hausman to incorporating moral commitments in models of rational choice. While the latter use a utility function that represents the agent's preferences, the former uses choiceworthiness functions that represent the evaluations of the moral theories in question. There is a difference between the choiceworthiness of an option according to a moral theory, and the desirability of an option to an agent, on the supposition that a particular theory is true. The question, "*if this theory were true, how much value would it assign to this option?*" is not the same as the question, "*if this theory were true, how much would I desire this option?*"

Theories of rational choice aim to model agents' preferences or desires with functions that assign values to options. If an agent is morally conscientious, then these values may track the values of the theory that they are considering. But the values nonetheless represent that agent's preferences. If an agent is less than fully conscientious, the moral theory may still play a crucial role in influencing their preferences. But this does not mean that the decision procedure that we adopt must take the values directly from the theory in question as its input. This is precisely what MEC does and it is this feature that lies at the root of the aforementioned problems of limited scope, intertheoretic comparisons and swamping by fanatical theories.

The problem of limited scope is that not all theories can be accommodated in MEC, since not all theories can be represented in the required numerical manner. Some theories provide no values whatsoever, and of those that do, not all have the structure required to calculate expected choiceworthiness. The attempt to use values that represent the choiceworthiness of options according to theories is directly responsible for this problem, since no such values are available for some theories.

The problem of intertheoretic comparisons is that different theories have different choiceworthiness functions and these cannot always be compared, either because they refer to different kinds of quantity (the Reference Problem) or because their values are unique only up to positive affine transformations and there is no way of fixing them on

the same scale (the Scale Problem). The attempt to use values that represent the choiceworthiness of options according to theories is directly responsible for this problem, since these values are not always comparable.

The problem of swamping by fanatical theories is that we can conceive of a theory $T_j$ such that $cT_j$ provides values that swamp the expected choiceworthiness calculation, even when assigned very low credence. The attempt to use values that represent the choiceworthiness of options according to theories is directly responsible for this problem, since these values are sometimes astronomically large, but MEC is required to incorporate them.

Thus, all of these problems stem from features of the choiceworthiness values that are taken as inputs to calculations of expected choiceworthiness: the theories under consideration either fail to provide any values at all, fail to provide values on an interval scale, fail to provide values that are intertheoretically comparable, or provide values that are too large to be handled sensibly.

In sum, there are two conclusions to be drawn from this section. The first is that MEC is only applicable for agents whose utility functions perfectly track the choiceworthiness functions that are taken to represent the evaluations of options according to the theories under consideration, i.e., those who are completely morally conscientious. Therefore, it is either unrealistic or overly demanding. The second is that the way in which this procedure takes these values as its input underlies three of the major problems facing the use of MEC under moral uncertainty. I will now show how, by finding a role for agent's own preferences in choice under moral uncertainty, we can save the Analogical View from the three problems facing MEC and generalise it to accommodate agents who exhibit less than complete moral conscientiousness.

## 5      The Cure

I have suggested that the problems facing MEC arise from the practice of assigning values to options according to choiceworthiness functions that represent the answer to the question, "*if this theory were true, how much value would it assign to this option?*" However, this is not the only way to instantiate the Analogical View. Instead, I propose that agents assign

54

numbers to options with a value function that represents the preferences that they would have if a given moral theory were true. That is, a preference ordering that is arrived at through a process of compromise between their self-interest and the recommendations of the moral theory under consideration, along with any other considerations they deem to be relevant to the choice at hand.

From this point onwards, much of the framework of MEC can be preserved. We can still represent a decision as comprising options $A_1$ to $A_m$, moral theories $T_1$ to $T_n$, a probability function $P$ representing the agent's credence in those theories, and some values assigned to the options under each theory. The difference lies in the way in which these values are assigned. Rather than considering only how choiceworthy a given theory would suggest each option is, we can instead consider how desirable the agent would find each option if that the theory was true. Formally, we simply replace the set of choiceworthiness functions $cT_1$ to $cT_n$ with a set of value functions $vT_1$ to $vT_n$ that assign numbers to options representing the value of that option to the agent in question, on the supposition that the theory in question is true. Everything else can remain unchanged, so the expected value to be maximised is given by:

$$EV(A_i) = \sum_{j=1}^{n} vT_j(A_i) \cdot P(T_j)$$

To be clear, these value functions should be taken to represent a real psychological quantity: desirability to the agent in question. This proposal, therefore, falls firmly under the umbrella of mentalist approaches to rational choice. This should not come as any surprise, since MEC itself is already a far stronger claim than the behaviouristic interpretation of normative EUT. In fact, on the behaviourist interpretation of EUT, moral uncertainty would pose no particular challenge for rational choice, since this account does not require agents to assign values to alternatives at all, but merely to have preferences that satisfy the axioms of a representation theorem. If anything, this might be seen as a limitation of behaviourism, since "*have transitive preferences*" is an utterly unhelpful response to the question, "*how should I act under moral uncertainty?*"

This approach shares much of its formal machinery with MEC. Like MEC, it is still expectational in form. Like MEC, it still recommends a maximising approach. Like MEC,

it still treats moral uncertainty as roughly equivalent to empirical uncertainty. It might seem as though the decision procedure I am proposing is indistinguishable from the account that it is supposed to replace. However, there are a number of important points of difference in terms of interpretation and application between these siblings in the Analogical family. Most importantly, as I have already noted, the values assigned to options represent something different: not choiceworthiness according to the moral theory at hand, but desirability to the decision-maker on the supposition that the theory is true. This may be filled out in much the way that Sen suggested we deliberate on choices in light of our self-interested preferences and moral commitments.[69] In a choice under moral uncertainty, an agent would undertake a kind of hypothetical reasoning, taking each theory in turn and considering what their all-things-considered preferences would be if that theory were true; what compromise they would be willing to make between their self-interest and their moral commitments.

For example, recall the case described above of the philanthropic lottery winner, choosing what to do with their windfall—donate to Malaria Consortium or give money to their family—while uncertain about their moral obligations to their nearest and dearest. First, they imagine that the utilitarian theory $T_1$ is true, that what really matters is the overall balance of happiness over unhappiness in the world. Just as Sen suggests, they make some compromise between their moral commitment and other considerations, including self-interest, to arrive at the values that they would assign to the options. Next, they move on to consider the family oriented $T_2$. Again, they weigh up their self-interest and moral commitments under this moral supposition and assign values to the options. This process can fill out the decision matrix from which expected value can be calculated to guide the choice of what to do with the winnings.

Another difference between MEC and the current proposal is that the former is only applicable to agents who are completely morally conscientious, but the latter neither assumes nor requires any particular degree of moral conscientiousness. When an agent is imagining that a given moral theory is true, they may be concerned only with morality, as MEC assumes, but they may just as well hold a weaker moral commitment, reaching a different compromise with self-interest and assigning different values accordingly. The value functions that represent our lottery winner's preferences may well perfectly align with

---

[69] Sen, "Rational Fools: A Critique of the Behavioural Foundations of Economic Theory."

choiceworthiness functions that represent the competing moral theories. But they may just as well diverge quite significantly, if the recommendations of these theories are not all that matter to this individual's overall evaluation of the options. Therefore, while MEC is only applicable for fully morally conscientious agents, the current proposal has no such restriction. The relaxation of the requirement for complete moral conscientiousness will also be relevant when considering the problems that undermine MEC, to which we will turn our attention next.

## 5.1    Incorporating Valueless Theories

The first problem for MEC was that its scope is limited to those choices in which an agent's degrees of belief are divided only between theories that can be represented by values on an interval scale. This excludes a number of prominent approaches to moral thought, to which many people will give at least some credence. This is because MEC relies on the theories themselves to provide the right kind of values and if a given theory produces no such values, then it cannot be incorporated into MEC. However, under the current proposal, there is nothing to restrict the scope in this way. No matter what structure a moral theory takes, an agent can imagine that it is true and then evaluate alternatives accordingly. Suppose I am considering whether or not to tell my friend the painful truth about their singing voice, but am morally uncertain, giving some credence to the view that I should just do whatever makes my friend happiest overall and another view according to which there is an absolute moral prohibition against dishonesty. Even though the latter view provides no ordering or value assignment—it simply says that it is always wrong to lie—I can entertain the supposition that this deontological principle is correct, then imagine how desirable the options would be, all things considered, and assign values accordingly. In fact, something like this kind of reasoning takes place whenever someone makes a judgement about the desirability of an option that is influenced by a range of different considerations that includes their moral commitments. We often hear people say things like, "*I know that eating meat is wrong, but roast lamb is just so delicious.*" Even though this moral view only offers a simple proscription on eating meat, the all-things-considered desirability of the options can be compared to produce an ordering and assign values.

This sort of example might seem problematic to those who believe in the motivational internalism of moral judgements. This is the view that moral judgements are necessarily or

57

intrinsically motivating, such that if one judges something to be right, then one is necessarily motivated to pursue it. On closer inspection, however, there is no tension between this view and the sort of reasoning I describe. The presence of motivation to act, whether or not it arises necessarily from a moral judgement, does not preclude the agent in question from having preferences over the alternatives under consideration. If motivational internalism is correct, then moral judgements necessarily shape or constrain our preferences, by pushing the right options towards the top of the ordering. But this poses no problem for the suggestion that agents can consider moral theories that do not value or order options, and come up with the preferences that they would have if such a theory were true.

Alternatively, one might object that the move from choiceworthiness functions that represent moral theories to value functions that represent desirability will not help, since preference orderings themselves might lack the requisite features. For example, one might follow Larry Temkin and doubt that rational agents' preferences are always transitive, a necessary condition for representing preferences numerically. [70] This would indeed pose a problem for my proposal. However, this is no more of a problem for this procedure than for the orthodox accounts of EUT that it is an attempt to emulate. The central claim of the Analogical View is that *if* maximisation of expected value is the correct theory of rational choice under empirical uncertainty, *then* it is the correct theory of rational choice under moral uncertainty. This conditional cannot be rendered false if we deny the antecedent.

As a matter of fact, the current proposal is, if anything, a better embodiment of the spirit of the Analogical View than MEC. Advocates of MEC claim that the correct theory of rational choice under moral uncertainty should emulate the correct theory of rational choice under empirical uncertainty. It is odd, therefore, that they should devise a decision procedure based on the choiceworthiness functions of moral theories, rather than the utility functions that represent agents' desires. Confining their proposal to "*morally conscientious agents*" goes some way to explaining this, since such agents would, by definition, have utility functions that would track the choiceworthiness functions of the theory under consideration. But recall that it is rationally permissible for an agent to care about things

---

[70] Larry S. Temkin, *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning* (Oxford: Oxford University Press, 2012).

other than just morality. Incorporating agents' own preferences presents a more generalisable theory and goes further towards providing an account of choice under moral uncertainty that is analogical to choice under empirical uncertainty.

## 5.2    Allowing Intertheoretic Comparisons

The second challenge facing MEC was the Problem of Intertheoretic Comparisons. That is, comparisons of value between different theories are required for calculations of expected choiceworthiness, but these comparisons are not possible because the theories' choiceworthiness functions represent different kinds of quantity—the Reference Problem—and are not fixed to the same scale—the Scale Problem. It may at first appear as though, since the current proposal also uses different value functions in a single calculation of expected value, it will also suffer from a similar issue with comparability. However, it is worth considering the Reference and Scale problems in turn to see if there really is any reason to doubt the possibility of value comparisons here.

Firstly, the Reference Problem does not apply to the current proposal, since the different value functions are taken to represent the same quantity, namely the desirability of the options to the agent. We are not trying to compare utils and rights, like apples and oranges, but rather an agent's preferences if one theory were true with their preferences if a different theory were true. Since the current proposal works with value functions that refer the same entity, the Reference Problem is easily avoided. To be sure, there may be some options, under some theories, that one is unable to compare with other such option-theory pairs, even in terms of their own desires or preferences. But of course, this is nothing new. One may equally be unable to compare the desirability of going for a walk if it is sunny, with that of staying in with a book if it rains. There is a great deal of literature concerned with the sorts of choices in which one option is neither better than, worse than, nor equal to another. [71] I discuss just such choices at length in Chapter 4. That this poses a challenge to rational decision makers is not unique to choices under moral uncertainty and, on my account, does not arise from attempting to compare quantities of different kinds of value.

---

[71] See, for example: John Broome, "Is Incommensurability Vagueness?," in *Incommensurability, Incomparability, and Practical Reason*, ed. Ruth Chang (Cambridge: Harvard University Press, 1997); Ruth Chang, "Hard Choices," *Journal of the American Philosophical Association* 3, no. 1 (2017): 1–21; Joseph Raz, *The Morality of Freedom* (Clarendon Press, 1986).

Thus, the Reference Problem does not undermine this approach to choice under moral uncertainty.

Next, the Scale Problem. Assuming these value functions are measurable on an interval scale, they are still unique only up to positive affine transformations, so we still face the challenge of knowing which value assignments to compare. However, given that they are taken to represent the same entity, we might be able to normalise the different value functions onto a common scale in order to make sense of the required value comparisons. All we need in order to be able to do this is some fixed points that can be used to anchor the scale of different value functions. This role can be played by any two options whose desirability to the agent in question does not depend on which moral theory is true. Given that the value assigned to them does not depend on the moral theory under consideration, any two such options should be assigned the same value by each value function. Therefore, we can either see that different value functions are on the same scale already, if they assign the same values to the theory-invariant options, or we can transform the functions onto a common scale such that these fixed points are given the same values. The values assigned to all other options should then fall into place, with the appropriate value relative to the morally neutral options, thus fixing the different value functions on the same scale.

This is simplest when there are two options that are equally desirable, no matter which theory is true. However, the same result can be achieved indirectly so long as each pair of value functions can be connected by a series of fixed points. For example, even if there are no two options whose desirability is the same under $T_1$ and $T_2$, as long as there are two options that are equally desirable under $T_1$ and $T_3$ and a different two options that are equally desirable under $T_2$ and $T_3$, then the value functions representing desirability under these three different moral theories can be normalised onto the same scale as $vT_3$, thus rendering comparisons between $vT_1$ and $vT_2$ possible.

The question, then, is whether any such fixed points can be found. And this, ultimately, depends on the particular desires of the agent in question. It is conceivable that, if a certain theory were true, it would completely change one's outlook and the evaluation of all options would be different to those under the supposition that a different moral theory was correct. However, it is plausible to think that there are often some options whose desirability to a particular agent is independent of which moral theory is true. In many

choice situations there are morally neutral options: options about which the different moral theories either agree or say nothing at all. Given that there is no disagreement between the moral theories in their evaluation of these options, the desirability of the options to the agent is likely to be independent of which moral theory being considered.

This line of reasoning echoes that of Jacob Ross [72] and Andrew Sepielli. [73] For example, Ross offers the following observation:

> *"There are some cases in which the task of commensurating two alternative ethical theories will be greatly facilitated. These are cases in which, for some pair of options, we know that the difference between their values is the same according to both ethical theories. This will obtain whenever there is a pair of options that does not differ in relation to any of those factors about which the two theories disagree."* [74]

The difference here is that, unlike Ross and Sepielli, I do not need to show that different theories agree in precisely the way required to commensurate their value functions. I only need to show that an agent's *own* evaluation of the options is independent of which moral theory is true.

Let us return to the example of the philanthropic lottery winner to make this clear. Suppose that, in addition to giving the money to Malaria Consortium or to their family members, this person also considered the options of spending the money on a Ferrari, or rejecting the winnings altogether. As before, they take each theory in turn and consider how desirable the options would be to them if that theory were true. If agent-neutral theory $T_1$ were true, then giving to family members would be considered preferable to buying a Ferrari, because of this agent's philanthropic motives. Buying a Ferrari would be preferred to rejecting the winnings altogether, because this person likes luxury sports cars. But giving to Malaria Consortium would be considered by far the best option, because of the amount of suffering that could be alleviated by donating, the agent-neutral value of alleviating suffering according to $T_1$, and the agent's moral commitment. This could be represented by the value function $vT_1$ such that $vT_1$(Malaria Consortium)=1000, $vT_1$(Family)=10,

---

[72] Ross, "Rejecting Ethical Deflationism."
[73] Sepielli, "What To Do When You Don't Know What To Do"; Andrew Sepielli, "Decision-Making Under Moral Uncertainty," in *Routledge Handbook on Moral Epistemology*, ed. Aaron Zimmerman, Karen Jones, and Mark Timmons (New York: Routledge, 2019), 508–21.
[74] Ross, "Rejecting Ethical Deflationism" p.764.

$vT_1$(Ferrari)=4, and $vT_1$(Reject)=1. Next, they consider agent-relative and family-oriented theory $T_2$ and, through similar reasoning, come up with something like the following: $vT_2$(Malaria Consortium)=40, $vT_2$(Family)=60, $vT_2$(Ferrari)=20, and $vT_2$(Reject)=5.

The difference between $T_1$ and $T_2$ is the weighting of benefits given to the agent's nearest and dearest and this is not relevant to the new options being considered. Neither option achieves the agent's aim of using the money to help other people and the agent's preference for buying a Ferrari over rejecting the winnings is motivated only by self-interest. They are morally neutral options and their desirability does not depend on which moral theory is true. Therefore, they should be assigned the same value under the two theories. We can use these options as fixed points around which to normalise the values of the options. In this case, we can convert $vT_2$ to $vT_2^*$ by dividing all of the values by 5, so that the morally neutral options have the same values under either theory. This is shown in Table 6 below.

**Table 6.** Normalised value functions

|  | $vT_1$ | $vT_2$ | $vT_2^*$ |
|---|---|---|---|
| Give to Malaria Consortium | 1000 | 40 | 8 |
| Give to family | 10 | 60 | 12 |
| Buy Ferraris | 4 | 20 | 4 |
| Reject winnings | 1 | 5 | 1 |

Now, given that the morally neutral options have the same values under the different theories, we know that the two value functions $vT_1$ and $vT_2^*$ are fixed to the same scale. This means that we can meaningfully compare the values under these two theories in order to calculate the expected value of the options, given the agent's credence in the competing moral views. Crucially, we have only rescaled the original value functions by positive affine transformations, so the ratios of value difference between the options have been preserved. In other words, the difference between giving to their family and buying a Ferrari if $T_2$ were true is still twice the difference between giving to their family and giving to Malaria Consortium, and so on for the other ratios of value difference. Therefore, the normalised functions contain the same information as the originals, but we can now make sense of

comparisons of value between them, since they are defined relative to the morally neutral options.

The move from functions representing choiceworthiness according to a theory, to functions representing desirability to an agent means that the Problem of Intertheoretic Comparisons does not necessarily apply to choice under moral uncertainty. Whether the required comparisons can actually be made in practice, however, depends on whether or not there are some options whose desirability is independent of the moral theory under consideration and this, in turn, depends on the particular psychology of the agent in question. Note, however, that this isn't a challenge unique to choice under moral uncertainty; there is no guarantee that one can always commensurate the desirability of options in different states of the world under empirical uncertainty. If one can find equally desirable options to use as fixed points between different theoretical suppositions, then one can commensurate the value functions in order to make the comparisons required by the Analogical View.

## 5.3     Escape from Fanatical Theories

The last problem facing MEC is that calculations of expected choiceworthiness can be swamped by 'fanatical' moral theories, which assign such large values to some options that they dictate what the agent in question ought to do, even if that agent considers any such theories highly unlikely to be correct. This problem can be avoided on the current proposal by noting that one's commitment to morality may depend on the particular moral theory that is under consideration. Under MEC, we have to use the values that are produced by the theories in question. If one of these theories considers the matter at hand sufficiently high-stakes, then the choice will be dictated by that theory, even if it is considered highly unlikely. On the current proposal, however, we do not have to use whatever values are provided by the theory in question. Rather, we assign values based on the degree to which we would be swayed by a theory if it were true. Given that we are not rationally required to have any particular degree of commitment to moral considerations, we needn't assign values that are large enough to swamp this choice procedure, even when considering moral theories that make outlandish claims about the moral value of certain options. One's degree of commitment to one moral theory, if it were true, need not be the same as one's degree of commitment to a different moral theory if that were true instead. There are

certain moral theories of which one might reasonably say, "*if* that's *what morality says, then I don't want to be moral!*" This follows from what we have already established about the nature of moral commitment and rational choice, but it also seems to make sense at an intuitive level. It does not seem implausible to think that if we somehow learnt that morality was incredibly demanding we might only be willing to make limited sacrifices in the trade off between morality and self-interest. Moreover, doing so would not be instrumentally irrational. Being concerned with matters other than morality might well be morally criticisable, but is surely not legitimate grounds for claims of irrationality.

To illustrate this point, think of two differing accounts of the morality of giving to charity. First, the commonsense view according to which donating to charity is good, but not donating to charity is not wrong. Those who give to charity may be praised for their generosity, but this is over and above what is required and it is permissible for individuals to choose to spend their hard-earned money on new clothes and holidays instead. Contrast this with Peter Singer's claim that, "*we ought to be preventing as much suffering as we can without sacrificing something else of comparable moral importance*" [75] and the consequence of this view that we should give away our wealth "*up to the point at which by giving more one would begin to cause serious suffering for oneself and one's dependents.*" [76] On Singer's view the stakes are very high. Not giving to charity is morally equivalent to choosing not to save a child from drowning in a pond, so as not to get one's clothes wet. Given how much is claimed to be at stake, we should take this view seriously; the risk of doing something seriously morally wrong should have some affect on a choice, even if this is not considered the single moral theory most likely to be correct. However, for an agent who feels almost certain that Singer's argument is unsound, it seems absurd to say that rationality requires that they should act as his conclusion demands. The problem for MEC is that it gets swamped by high stakes theories such as this.

On the current proposal, however, this swamping can be avoided, because there is no need for one's value function to exactly track choiceworthiness according to Singer's high-stakes theory. When considering a choice about charitable giving, one might say that, if the commonsense view is correct account then fine, I will give a proportion of my income to charity. If Singer is right, on the other hand, I will give more than I would have done if the

[75] Peter Singer, "Famine, Affluence, and Morality," *Philosophy and Public Affairs* 1, no. 3 (1972): p.238.
[76] Singer, p.234

commonsense view was correct, but I will simply never choose to lower my own standard of living to the degree that he demands. That theory is just too demanding for me to comply with completely. There seems to be no reason why a statement like this should be considered irrational. One might criticise this approach on moral grounds—morally speaking, one should be willing to do whatever morality says—but it is unclear why we should judge there to be anything irrational about this sort of reasoning. Morality is not the only consideration that rational agents care about. People have their own personal interests in addition to their moral commitments and it is not irrational for them to make trade offs between these competing considerations. There is nothing irrational about one's commitment to a moral theory being sensitive to what that theory says.

That said, the account I have offered is still sensitive to stakes to some degree. Recall that there is something of a double-edged sword, with sensitivity to stakes on one side and swamping by fanatical theories on the other. The current proposal captures the advantages of the former without succumbing to the exorbitant demands of the latter. As the previous example illustrated, one might be willing to give *more* if Singer is right, just not *as much more* as he demands. Such an agent is sensitive to how much is at stake, but not sensitive enough to be swamped by fanatical theories.

This sort of reasoning might be subject to another objection from motivational internalism. For the motivational internalist, if a high-stakes theory were true, then rational agents would have a corresponding motivation to pursue the action that it recommends. Swamping by such theories could not so easily be avoided by choosing on the basis of subjective desirability rather than moral choiceworthiness, since the former will track the latter, according to motivational internalism. In response to this point, it will be instructive to distinguish between two forms of motivational internalism. Weak motivational internalism holds that moral judgements necessarily provide some degree of motivation. Judging that an action is right always provides *some* reason to pursue it, but this can be overridden by countervailing reasons. This form of motivational internalism in consistent with the proposal for avoiding swamping by fanatical theories. Although weak motivational internalism requires that an option that is judged to be morally right is pushed towards the top of one's preference ordering, the strength of motivation provided by a moral judgement need not be equivalent to exactly how much is at stake according to that view.

65

Strong motivational internalism, on the other hand, holds that moral judgements necessarily provide *overriding* motivation to act as they demand. This is a view that J. L. Mackie ascribed to Plato: of Plato's Forms, especially the Form of the Good, Mackie says, "*just knowing them or "seeing" them will not merely tell men what to do but will ensure that they do it, overruling any contrary inclinations.*" [77] It does seem true that strong motivational internalism, when coupled with the current approach, results in swamping. This means that there is conflict between three claims: (i) strong motivational internalism, (ii) the Analogical View, and (iii) the view that swamping is an unacceptable feature of an account of choice under moral uncertainty. But it is not immediately clear why the presence of this conflict should lead us to reject the Analogical View, rather than one of these other claims. This argument could just as well be used to show that strong motivational internalism entails an implausible conclusion, or that we should reconsider our aversion to swamping. In fact, strong motivational internalism and the assertion that swamping is a problem seem like the oddest bedfellows of the three. If one believes that moral judgements necessarily provide indefeasible motivation, then it seems fitting to think that the possibility of a high-stakes moral view should determine how one acts. In summary, strong motivational internalism, the Analogical View and the judgement that swamping is a problem are mutually inconsistent. But, absent further argument, this does not show that the Analogical View, rather than one of these other claims, should be abandoned.

So, just like the other problems before it, swamping by fanatical theories can be avoided my assigning values to options according to all-things-considered desirability, rather than moral choiceworthiness.

## 6    Conclusion

The question, "*if this theory were true, how much value would it assign to this option?*" is not the same as the question, "*if this theory were true, how much would I desire this option?*" It is this method of value ascription via choiceworthiness according to theories, rather than desirability under the supposition that a theory is true, that I have shown is the root cause of three major problems for MEC. The current proposal can avoid these problems, while

---

[77] John Leslie Mackie, *Ethics: Inventing Right and Wrong* (New York: Penguin Books, 1977) p.23.

preserving MEC's advantageous features and thereby presents a superior instantiation of the Analogical View. Furthermore, it provides a role for agents' own preferences or desires that is conspicuous in its absence from MEC and generalises the Analogical View to agents who do not care only about morality. Therefore, we are able to treat moral uncertainty as roughly analogous to empirical uncertainty when modelling rational choice.

To some, the product of this paper is a philosophical machine for decision making. Input your values and probabilities and out pops a concrete answer to a decision problem. To others, however, the mechanistic structure of the proposed decision procedure, along with the numerical representation of values and degrees of belief, may imply an implausible degree of precision in our moral and rational deliberation. Even if one finds the formal machinery of decision theory an unappealing approach to real choices, however, it can provide an idealised framework within which to hammer out debates about, *inter alia*, the way in which our choices reflect our judgements about the likelihood of different moral views and the amount that is at stake according to them; about the way in which we can compare the judgements of different theories that we think could be true; and about how to deal with moral theories that seem to make judgements so extreme that they threaten dominate our choices. Even for those who won't calculate an expected value for as long as they live, this reasoning might provide the conceptual tools to think about choices that depend on uncertain moral beliefs.

# CHAPTER 3

## Attitudinal Ambivalence: Moral Uncertainty for Non-Cognitivists

### 1      Introduction

> *If we conceive that a thing, which is wont to affect us painfully, has any point of resemblance with another thing which is wont to affect us with an equally strong emotion of pleasure, we shall hate the first-named thing, and at the same time we shall love it. […] This disposition of the mind, which arises from two contrary emotions, is called vacillation; it stands to the emotions in the same relation as doubt does to the imagination.*
>
> Spinoza 1667: E3p17

It has been argued that non-cognitivist attitudes have insufficient internal structure to accommodate the features that common sense tells us are part of moral judgements. Whereas moral judgements vary in certitude, importance, and robustness, non-cognitive attitudes only possess two structural features: strength and stability over time. Therefore, moral judgements cannot be non-cognitive attitudes. [78]

Could the *"vacillation"* described by Spinoza reveal a feature of non-cognitive attitudes that is overlooked by this critique? [79] In this paper, I will argue that such internal conflict, or ambivalence, can play the role of moral doubt that is missing from conventional non-cognitivist accounts of moral judgements.

In addition to providing a defence of non-cognitivism, reinterpreting the phenomenon of moral doubt in terms of attitudinal ambivalence provides an alternative perspective to the

---

[78] Michael Smith, "Evaluation, Uncertainty and Motivation," *Ethical Theory and Moral Practice* 5, no. 3 (2002): 305–20.

[79] Benedict de Spinoza, "Ethics," in *The Collected Works of Spinoza Volume I*, ed. Edwin Curley, vol. 1 (Princeton: Princeton University Press, 1996).

recent literature on moral uncertainty. Ambivalence, as a fundamentally different state to uncertain belief, requires a fundamentally different response.

§2 will present the problem, retracing Michael Smith's argument that non-cognitivism cannot accommodate all three structural features of moral judgements. §3 will introduce and characterise the notion of ambivalence. §4 will demonstrate how ambivalence can play the role of certitude and thus provide a defence of non-cognitivism, while §5 will anticipate and respond to some potential objections to this defence.

## 2      The Problem: Non-cognitivism and Moral Uncertainty

What should we do when we do not know what we should do? There is a growing philosophical literature built on attempts to answer this question and others like it. A number of philosophers have argued for various norms that we should follow when uncertain about moral facts. [80] Others have argued that no such subjective norms exist, that what we should do is in fact what the true moral theory says we should do, regardless of our beliefs. [81]

The use of terms such as 'facts', 'belief', and 'true' appear to betray an underlying presupposition in favour of metaethical realism. Indeed, debates on the topic of moral uncertainty and the very question that defines it seem to depend on assumptions about the nature of morality that are by no means uncontroversial. Non-cognitivists, for example, would deny that any of these three terms accurately describe moral judgements. This is a form of moral anti-realism; its proponents deny the existence of objective moral facts. Rather than asserting that all moral beliefs are mistaken, however, non-cognitivists state that moral judgements are not beliefs at all and are neither true nor false. Moral

---

[80] Lockhart, *Moral Uncertainty and Its Consequences*; Ross, "Rejecting Ethical Deflationism"; Sepielli, "What To Do When You Don't Know What To Do"; Sepielli, "'Along an Imperfectly-Lighted Path': Practical Rationality and Normative Uncertainty"; Gustafsson and Torpman, "In Defence of My Favourite Theory"; MacAskill and Ord, "Why Maximize Expected Choice-Worthiness?"; MacAskill, Bykvist, and Ord, *Moral Uncertainty*.

[81] Harman, "The Irrelevance of Moral Uncertainty"; Weatherson, "Running Risks Morally"; Brian Weatherson, *Normative Externalism* (Oxford: Oxford University Press, 2019).

judgements, according to the non-cognitivist, are some variety of conative attitude, more akin to desire than belief.

Thus, non-cognitivism can be characterised as comprising three central claims:

1. **Moral antirealism:** There are no objective moral facts.

2. **Psychological non-cognitivism:**

    i.    Moral judgements are not beliefs.

    ii.   Moral judgements are some variety of conative attitude.

3. **Semantic non-factualism:** moral sentences are not truth-apt; they take no truth values.

This conception of non-cognitivism may be contested by quasi-realists, such as Simon Blackburn and Allan Gibbard, who adopt a minimalist theory of truth to claim that terms like 'fact', 'belief', and 'true' can be applied to moral judgements, even though they are not representational. [82] Indeed, there has been extensive debate about whether quasi-realism is equipped to make sense of self-doubt and moral fallibility. [83] However, I will set these accounts to one side for now and focus on non-cognitivism as characterised above, since quasi-realism is sufficiently distinctive to require its own, separate analysis.

Because uncertainty is a measure of degrees of belief and non-cognitivism holds that moral statements are not beliefs, discussion of moral uncertainty per se presupposes that non-cognitivism is false. However, might not this phenomenon be redescribed in a way that is consistent with non-cognitivism? Is there some gradable feature of non-cognitive attitudes

---

[82] Simon Blackburn, *Spreading the Word: Groundings in the Philosophy of Language* (Oxford: Clarendon Press, 1985); Simon Blackburn, *Essays in Quasi-Realism* (Oxford University Press, 1993); Allan Gibbard, *Thinking How to Live* (Harvard University Press, 2003).

[83] Andy Egan, "Quasi-Realism and Fundamental Moral Error," *Australasian Journal of Philosophy* 85, no. 2 (2007): 205–19; Simon Blackburn, "Truth and A Priori Possibility; Egan's Charge Against Quasi-Realism," *Australasian Journal of Philosophy* 87, no. 2 (2009): 201–13; Sebastian Köhler, "What Is the Problem with Fundamental Moral Error?," *Australasian Journal of Philosophy* 93, no. 1 (2015): 161–65; Bob Beddor, "Fallibility for Expressivists," *Australasian Journal of Philosophy* 98, no. 4 (2020): 763–77; Garrett Lam, "Making Quasi-Realists Admit of Fundamental Moral Fallibility," *Australasian Journal of Philosophy* 98, no. 2 (2020): 294–303; Stina Björkholm, Krister Bykvist, and Jonas Olson, "Quasi-Realism and Normative Certitude," *Synthese* 198 (2021): 794–99, https://doi.org/10.1007/s11229-020-02553-y.

that can stand in for moral uncertainty for those who deny that moral statements express beliefs? Michael Smith has argued that there is not. [84]

Smith argues that moral judgements have three structural features. The first is certitude. This is the degree of confidence one has in a judgement. For example, I feel more confident that it would be right to save a drowning child than that I should donate most of my wealth to overseas development charities. This is the feature that has gained much recent attention under the banner of normative or moral uncertainty. I shall refer to the phenomenon of low moral certitude as moral doubt.

The second is importance. This is a measure of how desirable or undesirable, good or bad, right or wrong, one judges something to be. Importance is distinct from certitude. For example, I am equally confident that lying is wrong and that torturing an innocent person is wrong, but I judge torture to be much worse than lying. These two judgements are equal in certitude, but they differ in importance.

Lastly, we have robustness. This is the stability of moral judgements over time, as an individual reflects and receives new information. Whereas certitude and importance are synchronic, robustness is diachronic: it is a measure of change through time. While Spinoza's use of the term "*vacillation*" refers to what I will call ambivalence, modern usage of the term might be thought to refer to a lack of robustness. Note that Smith defines robustness as the temporal stability of certitude only, but there may also be changes in importance over time. For example, as I learn about the various successes of non-violent protests in history and reflect on current political injustices, I may come to judge political activism as more important than I did previously, *as well as* feeling more confident in this judgment.

Any acceptable account of the nature of moral judgements must adequately capture certitude, importance, and robustness. The central claim in Smith's argument is that cognitivist metaethical theories can accommodate these three features, but that non-

[84] Smith, "Evaluation, Uncertainty and Motivation."

cognitivism cannot. This point can be made clear by comparing the archetypal cognitivist and non-cognitivist interpretations of a typical moral judgement.

Suppose an individual X makes the moral judgement that an action φ is morally right. The cognitivist will interpret this as follows:

*"X believes that φ is right."*

This judgement is the belief that φ has the property of rightness. There are two features that can vary independently: the degree of belief and the degree of rightness. The former captures the certitude of X's judgement and the latter captures its importance. The stability of these features through time gives the judgement's robustness. Thus, a cognitivist interpretation of moral judgements captures all three structural features.

The non-cognitivist interpretation, on the other hand, is something like:

*"X desires that φ."*

For simplicity, we will proceed with desire as the non-cognitive attitude in question, but note that there are numerous instantiations of non-cognitivism and this discussion applies to non-cognitivism broadly construed. Desire could be replaced with emotive responses,[85] prescriptions,[86] approval,[87] intentions or commitment to plans,[88] and the analysis herein should remain intact.

Rather than a property of φ and a belief about this property, on this interpretation there is only one variable component: a desire. Therefore, this desire alone will have to capture the certitude, importance, and robustness of X's judgement. However, as Smith points out, it

---

[85] W H F Barnes, "A Suggestion about Value," *Analysis* 1, no. 3 (1934): 45–46; Charles Leslie Stevenson, "The Emotive Meaning of Ethical Terms," *Mind* 46, no. 181 (1937): 14–31; Alfred Jules Ayer, *Language, Truth, and Logic* (New York: Dover Publications, 1952).

[86] Rudolf Carnap, *The Logical Syntax of Languae* (London: Routledge, 1937); Richard Mervyn Hare, *The Language of Morals* (Oxford: Clarendon Press, 1952).

[87] Frank Jackson, "Non-Cognitivism, Validity and Conditionals," in *Dale Jamieson (Ed.) Singer and His Critics* (Oxford: Blackwell, 1999), 18–37; Michael Ridge, "Ecumenical Expressivism: Finessing Frege," *Ethics* 116, no. 2 (2006): 302–36; Mark Schroeder, *Being For* (Oxford: Oxford University Press, 2008); Mark Schroeder, "Expression for Expressivists," *Philosophy and Phenomenological Research* 76, no. 1 (2008): 86–116.

[88] Allan Gibbard, *Wise Choices, Apt Feelings: A Theory of Normative Judgment* (Harvard University Press, 1990); Gibbard, *Thinking How to Live*.

seems that desires have only two gradable features: strength and the stability through time. The latter clearly represents robustness, but the strength of a desire cannot be equivalent to both the certitude and the importance of a moral judgement. For example, if my judgement, "*telling the truth is good*", is simply a desire for truth-telling I may desire this more or less strongly. But common sense tells us that there is a meaningful difference between a confident judgement that telling the truth is slightly good and a less confident judgement that telling the truth is extremely good. By only varying in strength, desires alone cannot capture this difference. In Smith's words:

> "*The problem that looms here, then, is that non-cognitivists seem not to have the resources to accommodate all three features of evaluative judgement that we commonsensically ascribe to them. They can accommodate either importance and robustness, or certitude and robustness, but not all of importance, certitude, and robustness.*" [89]

In summary, the argument goes as follows:

> **P1** Moral judgements have three structural features.
>
> **P2** Non-cognitive attitudes have only two structural features.
>
> **C** Moral judgements are not non-cognitive features.

This is presented as a critique of non-cognitivism. But note that a committed non-cognitivist may well perform a Moorean shift, inverting this argument to deny that moral judgements vary in both importance and certitude. If we take the strength of a non-cognitive attitude to capture the importance of a moral judgement, then the asymmetry between desire-like attitudes and moral judgements appears to show that we cannot have both non-cognitivism *and* moral doubt.

There have been a number of attempts to defend the compatibility of non-cognitivism and moral doubt in response to this argument. James Lenman and Michael Ridge have both

---

[89] Smith, "Evaluation, Uncertainty and Motivation." p.317.

proposed modified versions of non-cognitivism, introducing a belief component to moral judgements in addition to the desire-like attitude. [90] This is belief about the desires of some suitably idealised advisor. They therefore accept Smith's conclusion that non-cognitive attitudes alone cannot capture moral doubt, but add beliefs into the mix to do the job instead. While the strength of the ideal advisor's desires represents importance, the degree of belief about these desires represents certitude. However, Bykvist and Olson (2009) have shown that this move only allows these forms of ecumenical non-cognitivism to account for empirical uncertainty about what some descriptively specified ideal advisor would desire, rather than fundamental normative uncertainty. [91] Johan Eriksson and Ragnar Francén Olinder offer a similar response, subject to similar problems, as highlighted again by Bykvist and Olson. [92]

More recently, Andrew Sepielli [93] and Julia Staffel [94] have attempted to show that moral doubt can be accommodated without invoking cognitivist attitudes, by building upon Mark Schroeder's "*being for*" version of non-cognitivism. [95] According to this account, a moral judgement involves two non-cognitive attitudes: a conventional non-cognitivist attitude, such as blame or praise, and a second attitude, *being for*, which is directed at the first attitude. The *being for* interpretation of the earlier example of a moral judgement would be:

*"X is for praising for φ"*

By introducing two non-cognitive attitudes, rather than one, this account provides the structure to accommodate both importance and certitude. The extent to which X is for

[90] James Lenman, "Non-Cognitivism and the Dimensions of Evaluative Judgement," *Brown Electronic Article Review Service (BEARS)*, 2003; Michael Ridge, "Certitude, Importance, and Robustness for Non-Cognitivists," *Brown Electronic Article Review Service (BEARS)*, 2003; Ridge, "Ecumenical Expressivism: Finessing Frege."
[91] Krister Bykvist and Jonas Olson, "Expressivism and Moral Certitude," *The Philosophical Quarterly* 79, no. 6 (2009): 1225–48.
[92] John Eriksson and Ragnar Francén Olinder, "Non-Cognitivism and the Classification Account of Moral Uncertainty," *Australasian Journal of Philosophy* 94, no. 4 (2016): 719–35; Krister Bykvist and Jonas Olson, "Non-Cognitivism and Fundamental Moral Certitude: Reply to Eriksson and Francén Olinder," *Australasian Journal of Philosophy* 95, no. 4 (2017): 794–99.
[93] Sepielli, "Normative Uncertainty for Non-Cognitivists."
[94] Julia Staffel, "Expressivism, Normative Uncertainty, and Arguments for Probabilism," in *Oxford Studies in Epistemology, Volume 6*, ed. Tamar Szabó Gendler and John Hawthorne (Oxford: Oxford University Press, 2019), 161–89.
[95] Schroeder, *Being For*; Schroeder, "Expression for Expressivists."

blaming for φ captures the certitude of this evaluation and the degree of blame X is for captures the importance. This provides the orthogonality of certitude and importance that Smith claimed was missing: X could be strongly for weakly blaming for white lies and weakly for strongly blaming for blasphemy.

However, Bykvist and Olson have also highlighted a number of problems with this account. [96] For example, because of the close connection between an agent's motivation and the being for attitude, this account implausibly implies that an agent who lacks motivation, due to depression or general apathy, necessarily has lower certitude in their moral judgements. But we generally do not think that such motivational akrasia means that someone is less certain about moral matters, even if they are less likely to act on their moral judgements. A second problem with this approach is that it implies that the only reason one might not be strongly for blaming or praising for some action is low certitude. But sometimes we are not in favour of blaming for some action, not because of a lack of certitude in the judgement that it is wrong, but because blame is considered to be counterproductive, or because of some features of the blamer's standing in relation to the subject. [97]

This is a superficial overview of a substantial philosophical debate, but for now it will suffice to highlight the fact that the major responses to the challenge Smith posed for non-cognitivism face significant problems of their own. [98] These attempts to rescue non-cognitivism all share a basic strategy: they accept that Smith's argument undermines extant non-cognitivist theories and proceed to modify them with extra components, be they beliefs or further non-cognitive attitudes, to accommodate certitude. Here I want to take a different tack. Drawing on evidence from social psychology, I will show that P2 of the above argument is false and that even the most unsophisticated non-cognitive attitudes

---

[96] Krister Bykvist and Jonas Olson, "Against the Being For Account of Normative Certitude," *Journal of Ethics and Social Philosophy* 6, no. 2 (2012), https://doi.org/https://doi.org/10.26556/jesp.v6i2.63.
[97] This is a form of The Wrong Kind of Reasons problem. See: Roger Crisp, "Review of Kupperman, Value... and What Follows," *Philosophy* 75 (2000): 458–92; Daniel Jacobson, "Fitting Attitude Theories of Value (Stanford Encyclopedia of Philosophy/Spring 2011 Edition)," 2011, https://plato.stanford.edu/archives/spr2011/entries/fitting-attitude-theories/.
[98] For a summary of this debate and the problems facing the various extant attempts to accommodate moral uncertainty within non-cognitivism, see: MacAskill, Bykvist, and Ord, *Moral Uncertainty*, Chapter 7.

have sufficient structure to accommodate robustness, importance, and certitude. The missing ingredient was a feature of attitudes all along: ambivalence.

# 3 The Solution: Ambivalence

*Leontius the son of Aglaeon was coming up from the Piraeus, outside the North Wall but close to it, when he saw some corpses with the public executioner standing near by. On the one hand, he experienced the desire to see them, but at the same time he felt disgust and averted his gaze. For a while he struggled and kept his hands over his eyes, but finally he was overcome by desire; he opened his eyes wide, ran up to the corpses, and said, "There you are, you wretches! What a lovely sight! I hope you feel satisfied!"*

Plato, *Republic*: 439e–440a

Leontius faces a problem in deciding how to act: he concurrently wants to see the corpses and to avert his gaze. He has inconsistent, conflicting desires, which cause doubt and hesitation. In the end, he gives in to his morbid fascination and looks, but it is his state of mind before this point that is of interest here.

Recall P2 of Smith's argument: non-cognitive attitudes have only two structural features. These are the strength of an attitude and its stability over time. Stability is diachronic, so at a single moment non-cognitive attitudes have only a single structural feature: strength. How might the case of Leontius fit into this unidimensional account of non-cognitive attitudes? His attitude towards seeing the corpses appears to contain more complexity than merely its strength. This suggests that P2 is worth closer inspection.

To scrutinise an empirical claim about attitudes, such as P2, it seems appropriate to turn to the discipline of social psychology, historically branded as the description and explanation of attitudes. [99] For much of the last century, the scientific consensus seems to

---

[99] William Isaac Thomas and Florian Znaniecki, *The Polish Peasant in Europe and America: Monograph of an Immigrant Group* (Chicago: University of Illinois Press, 1918).

have been with Smith. In most psychological research, attitudes were considered unidimensional and standardly measured by the semantic differential technique, with positive and negative valence represented as opposite ends of the same scale. As Thompson, Zanna, and Griffin note, the very methods used to measure them "*construe attitudes as falling on some point along an evaluative dimension, assessed by such bipolar scales as: favourable-unfavorable, positive-negative, good-bad, or like-dislike.*" [100] So, if the attitude in question was desire and the object of this desire was a proposition *p*, we would have a strong desire that *p* at one end, a strong desire that ¬*p* at the other end, and indifference somewhere in the middle.

However, in 1972, K. J. Kaplan proposed a modification to the semantic differential technique. He claimed that the positive and negative components of attitudes should be measured separately, rather than on a single, bipolar scale. [101] On this method, the degree to which an individual desires that *p* would be measured on one scale, from no desire that *p* to strong desire that *p*, and the degree to which they desired that ¬*p* would be measured separately. In many simple cases of desire, this would yield the result that an individual has a desire that *p* and no desire that ¬*p*. In simple cases of indifference, this would yield the result that an individual has no desire for either. However, from its earliest uses, this modified semantic differential technique also revealed what Plato had been getting at more than two thousand years earlier: individuals can evaluate something as both positive and negative concurrently. They can both desire that *p* and desire that ¬*p*. They can be *ambivalent*.

The early models of attitudes, varying along a single dimension, seem unable adequately to represent Leontius's desires with respect to seeing the corpses. It would be misleading to suggest that he is somewhere in the middle, neither wanting to look nor wanting not to. In fact, the opposite is true: he holds both of these strong desires at once. There is a clear difference between a person like Leontius, who has a strong desire to do something *and* a strong desire not to do that same thing, and a person who has no strong feelings either

---

[100] Megan M Thompson, Mark P Zanna, and Dale W Griffin, "Let's Not Be Indifferent about (Attitudinal) Ambivalence," in *Attitude Strength: Antecedents and Consequences*, ed. R. E. Petty and J. A. Krosnick, vol. 4 (New York: Lawrence Erlbaum Associates, Inc., 1995), 361.
[101] Kalman J Kaplan, "On the Ambivalence-Indifference Problem in Attitude Theory and Measurement: A Suggested Modification of the Semantic Differential Technique.," *Psychological Bulletin* 77, no. 5 (1972): 361.

way. If we represent desires as varying along a single dimension, however, we cannot distinguish between these two states. Kaplan's proposed modification of the semantic differential technique was an attempt to provide a conceptual and methodological distinction between indifference and ambivalence in individuals' attitudes. This attempt succeeded and ambivalence has proven to be a rich vein of psychological research ever since.

There are, of course, some pitfalls in this sort of cross-disciplinary approach, because terms are not always taken to have the same meaning in different fields. For instance, the indifference that Kaplan refers to—the absence of any strong feeling either way—is different from the meaning of indifference within philosophy, where decision theorists take it to mean that neither of two options is strictly preferred to the other. In social psychology, one may be ambivalent, indifferent, or neither, but not both. Decision theoretic indifference, however, is entirely consistent with ambivalence if the opposing attitudes lead one to have no strict preference for one option over another. Similarly, attitudes are not conceived in exactly the same way in philosophy and psychology. Whereas an attitude in psychology is taken to be an evaluative appraisal of an object of almost any kind, including social groups, policy positions, abstract concepts or physical objects, attitudes in philosophy are taken to be the relation an individual bears to a proposition, and need not involve any psychological element whatsoever. As noted by Frank Ramsey, *"the beliefs which we hold most strongly are often accompanied by practically no feeling at all; no one feels strongly about things he takes for granted."* [102]

There is, however, an important lesson that can be learnt from the psychological study of ambivalence, which is relevant for the philosophical analysis of attitudes: that we can have a varying degree of internal conflict in our attitudes and that this affects the ways in which we think about and act on them. We simply require a definition of ambivalence that coheres with a philosophical understanding of the relevant concepts.

A desire provides an agent with at least *pro tanto* reason to act. My desire for ice cream provides me with some reason to seek ice cream out. Often, our desires give rise to reasons

---

[102] Ramsey, "Truth and Probability", p.65.

that cannot all be satisfied. It is this presence of competing reasons that I aim to capture with the label of 'ambivalence'. Reasons are competing if they support mutually exclusive options, or courses of action. This conception of ambivalence does not require a commitment to any metaphysically loaded notion of objective, normative reasons. The reasons at play here should be thought of as an agent's particular, contingent, motivating reasons. [103] These need not be 'good' reasons, nor must they reflect any moral, prudential, or even empirical truths beyond the simple fact that they motivate the agent in question.

Furthermore, reasons may apply not only to concrete courses of action, but also to mental attitudes themselves. Indeed, in so far as we take ourselves to have reasons for our desires, we may even be ambivalent with respect to which desires to have in the first place. The same can be said of other propositional attitudes for which we have reasons. An agent who has some reason to believe or desire that $p$, and some reason not to believe or desire that $p$, should be considered ambivalent even though these mental attitudes are not conventionally considered options of choice. Thus, for any two alternatives A and B, an agent is ambivalent iff:

i.   The agent has motivating reason for A.

ii.  The agent has motivating reason for B.

iii. If B then ¬A.

The mutual exclusivity of A and B may be necessary, due to some essentially inconsistent features of the options. This is most obvious when B simply is ¬A. Note that this allows for ambivalence in one's attitude towards a single object. Leontius has reasons for not looking at the bodies, he has reasons for looking at the bodies, and if he does look, he cannot not look. Thus, he is ambivalent, and his ambivalence regards a single object. Alternatively, A and B may only contingently clash, as when one is invited to two parties on the same evening, so cannot attend both. When this is so, it is possible that one may have competing reasons without being consciously aware of any ambivalence and lacking

[103] Joseph Raz, *Practical Reason and Norms* (Oxford: Oxford University Press, 1975); Michael Smith, *The Moral Problem* (Oxford: Blackwell, 1994); Derek Parfit and John Broome, "Reasons and Motivation," *Proceedings of the Aristotelian Society, Supplementary Volumes* 71 (1997): 99–146.

any feelings of internal conflict. These cases, known as potential ambivalence, can be safely set aside from the current discussion, because we are interested primarily in the sort of felt ambivalence that gives rise to feelings of tension and doubt.

If an agent is ambivalent in virtue of the reasons that are provided by their moral judgements, then they may aptly be described as morally ambivalent. Ambivalence may be due to conflict in desires, or due to conflicting reasons that arise from any other source. Thus, ambivalence is relevant for a wide range of non-cognitivist views concerning which attitudes comprise moral judgements. Of course, we may also be ambivalent in our reasons for action that play no part in our moral psychology. Buridan's ass is clearly ambivalent, torn between reasons for two equally delicious piles of hay, even though it makes no moral judgement about the conflicting options.

The degree of an agent's ambivalence is jointly determined by the balance and weight of their reasons. An agent is ambivalent to the degree that their conflicting reasons are strong and closely balanced. Thus: (1) the stronger conflicting reasons are, the greater the degree of ambivalence and (2) the more closely balanced conflicting reasons are, the greater the degree of ambivalence. Thus, an agent may be considered ambivalent when they have conflicting reasons, even if one set of reasons decisively outweighs the others to settle a choice. This may appear to clash somewhat with the every-day use of the term, which is normally reserved for cases in which someone has mixed feelings that are at least approximately equal in strength. However, the present definition is apt, since even when one set of reasons decisively outweighs the other, the outweighed reasons persist and are relevant to practical deliberation. As Martha Nussbaum notes, sometimes a person has conflicting desires, but is clear about which is the better choice by their own lights. [104] When an agent is ambivalent, they cannot avoid taking an action that they have at least some reason against, no matter what they choose. This may make a choice difficult, require careful deliberation and leave the nagging tug of residual reasons, even if an agent acts in accordance with the most compelling reasons, all things considered.

---

[104] Martha Nussbaum, "Aeschylus and Practical Conflict," *Ethics* 95, no. 2 (1985): 233–67.

The definition I have provided bears some important similarities and differences to a variety of philosophical conceptions of ambivalence and related concepts in a somewhat fragmented literature on the topic. One line of inquiry concerns the place of ambivalence in philosophical and scientific theories of emotion. In these discussions ambivalence is, quite naturally, characterised as the presence of conflicting or discordant emotions, although precisely what this conflict or discordance amounts to is often left unspecified. [105]

Another philosophical angle on ambivalence stems from the work of Harry Frankfurt and concerns the significance of ambivalence and the opposing state of wholeheartedness for individual wellbeing and the living of a fulfilled life. [106] Frankfurt's characterisation of ambivalence is complex and multifaceted, and includes incoherence in desire, inconsistency in preferences and not knowing what one really wants. For Frankfurt, however, the defining feature lies not in one's first order desires, but in one's second order attitudes towards them. If one identifies with a desire and wants to be moved to action by it, then it is their will. Ambivalence, then, is conflict in the second order attitudes that comprise a person's will. Frankfurt's central claim about ambivalence is that it is a bad thing for agents, "*a disease of the mind*" that we should strive to remedy in order to become wholehearted. Unless one is wholehearted, he argues, life will be lacking in meaning and satisfaction, guided by a will that is self-defeating and ultimately incoherent. Numerous philosophers have since argued that ambivalence is not inherently undesirable, but is the only reasonable response to a world in which our ultimate ends come into conflict with one another. [107] Richard Moran, for one, claims that:

> "*The person who discovers that two of his aims cannot be satisfied together does not yet have reason to abandon either of them. They may still be accurate reflections of what is*

---

[105] Patricia Greenspan, "A Case of Mixed Feelings: Ambivalence and the Logic of Emotion," in *Explaining Emotions*, ed. Amelie Oksenberg Rorty (Berkeley: University of California Press, 1980), 223–50; Kristján Kristjánsson, "The Trouble with Ambivalent Emotions," *Philosophy* 85, no. 4 (2010): 485–510; Bill Wringe, "Ambivalence for Cognitivists: A Lesson from Chrysippus?," *Thought: A Journal of Philosophy* 6, no. 3 (2017): 147–56.

[106] Harry G Frankfurt, *The Importance of What We Care About* (Cambridge University Press, 1988); Harry Frankfurt, *Necessity, Volition, and Love* (Cambridge University Press, 1999); Harry Frankfurt, *The Reasons of Love* (Princeton: Princeton University Press, 2004).

[107] Richard Moran, "Review Essay on 'The Reasons of Love,'" *Philosophy and Phenomenological Research* 74, no. 2 (2007): 463–75; Jennifer S Swindell, "Ambivalence," *Philosophical Explorations* 13, no. 1 (2010): 23–34; S D Feldman and A Hazlett, "In Defense of Ambivalence" (manuscript, n.d.).

*genuinely desirable and good in the world, and as such, he would be wrong to act as if the recognition of their incompatibility somehow discredited their worthiness to be cared about or pursued. [...] Ambivalence may be seen as a positive good, or part of a healthy relationship of the will to the complexity and ambiguity of the world it confronts."* [108]

Lastly, some philosophers have characterised ambivalence, sometimes termed internal or practical conflict, more simply as the holding of opposed desires. [109] That is, desires that are not jointly satisfiable. Others have cast doubt the conceptual possibility of desires that one knows are inconsistent and attempted to provide an alternative explanation of the observation that we often feel torn when forced to make a choice. [110]

These various conceptions of ambivalence will regularly coincide with one another and with the presence of competing reasons: the ambivalent attitudes of social psychology and mixed feelings of affective science and philosophy may well generate competing reasons, while the conflicting desires discussed in debates around practical conflict may be the product of, or give rise to, reasons for mutually exclusive options. Nonetheless, it is important not to conflate these closely related, but ultimately distinct conceptions of ambivalence and to avoid the mistake of thinking that what is true of one is automatically applicable to all. For the time being, I shall not explore these alternative conceptions of ambivalence further. I do not intend to take a stance on the related debates, nor to claim that the presence of competing reasons is the "correct" conception. Whatever one's view on the nature of and relations between emotions, desires, and other attitudes, it should at least be clear that the state of ambivalence as I have described it exists. I take it as an empirically and conceptually uncontroversial fact that agents often have motivating reasons for mutually exclusive options. My intention is to explore what this phenomenon can reveal about our moral judgements.

---

[108] Moran, "Review Essay on 'The Reasons of Love'", p.474
[109] Monika Betzler, "Sources of Practical Conflicts and Reasons for Regret," *Practical Conflicts: New Philosophical Essays*, 2004, 197–222; Hili Razinsky, "An Outline for Ambivalence of Value Judgment," *The Journal of Value Inquiry* 48, no. 3 (2014): 469–88.
[110] Donald Davidson, "How Is Weakness of The Will Possible?," in *Essays on Actions and Events* (Oxford: Clarendon Press, n.d.), 21–42; Frank Jackson, "Internal Conflicts in Desires and Morals," *American Philosophical Quarterly* 22, no. 2 (1985): 105–14, https://doi.org/10.2307/20014086; Derek Baker, "Ambivalent Desires and the Problem with Reduction," *Philosophical Studies* 150, no. 1 (2010): 37–47.

Perhaps there exist some attitudes, emotions or desires that are in some sense conflicting, but do not feature as reasons for agents. For example, the emotion of regret regarding a one-off event in the past appears not to provide any particular reason for an agent to do anything, but seems to conflict with feelings of satisfaction or pride about the same event. Similarly, conflicting desires for things that are entirely out of one's hands, such as the desire for both sun and rain, may seem to manifest ambivalence without providing reasons. [111] However, recall that we are considering a broad range of reasons here, including those for the adoption of mental attitudes. These emotions and desires may themselves provide competing reasons for adopting further mental attitudes, thereby satisfying the criteria for ambivalence so defined. In fact, I do not deny the possibility of some forms of internal conflict without competing reasons. This possibility is simply not relevant to the present discussion. We are looking for a feature of moral judgements that may contribute to our understanding of the nature of moral doubt and the presence of competing reasons may provide just such a feature. This account of ambivalence may not encompass all varieties of internal conflict, but mental states that do not provide reasons in the way that moral judgements do are not relevant to understanding the phenomenon of moral doubt.

Furthermore, any full account of ambivalence that is not based on reasons for mutually exclusive options will have to provide an explanation of what it means for attitudes, emotions, or desires to be opposed or conflicting. In some cases this is intuitively clear: like and dislike, love and hate, happy and sad. However, in others it is not so obvious. Many attitudes and emotions do not clearly exhibit polarity, while others seem to conflict without being directly opposed. What attitude is the opposite of regret? Are love and fear opposed? Defining ambivalence in terms of reasons has the practical advantages of clearly diagnosing the phenomenon without having to answer these questions and capturing conflict between mental states that might not be considered directly opposed.

This is a fairly general conception of ambivalence, lacking the precision of the various social psychological methods for quantifying ambivalence from the strength and closeness of conflicting attitudes. For the purposes of this paper, however, it will suffice to acknowledge that ambivalence exists as a result of the balance and weight of conflicting reasons, without the need for an exact measure. Recognising that agents can be ambivalent

---

[111] Jesse J Prinz, *Gut Reactions: A Perceptual Theory of Emotion* (Oxford: Oxford University Press, 2004).

in their desires, along with a range of other attitudes, shows that P2 of Smith's argument is false. Non-cognitive attitudes do not only vary in strength and robustness, but also in their degree of ambivalence. However, more is required in order to reconcile moral doubt and non-cognitivism: ambivalence in non-cognitive attitudes must be able to play the role of low certitude in moral judgements.

# 4    Playing the Role of Certitude

As in the case of Leontius, the presence of ambivalence may cause an agent to be unsure of how to act, making it difficult to know what to do and producing a form of practical doubt that is distinct from partial belief. Ambivalence poses a challenge when deciding what to do not because of a lack of information, like uncertainty, but because weighing competing reasons is difficult. In cases of moral ambivalence, this leads to doubt about what moral statements to assert, how to appraise the actions of others and, ultimately, what choices we make as morally conscientious individuals. The more ambivalent an agent's attitudes, the less confidence they may have in their moral judgements. Just as uncertainty gives rise to moral doubt according to cognitivism, ambivalence does so according to a non-cognitivist view.

Consider a classic case of what cognitivists would describe as moral uncertainty. In his 1946 lecture "*L'existentialisme est un humanisme*", Jean-Paul Sartre recounted the story of a student of his who, during World War II, was struggling to choose between leaving home to join the Free French Forces, or staying in France with his mother. [112] On the one hand, he yearned to fight the Nazis and avenge the death of his brother. On the other, he felt duty-bound to care for his mother, who would be devastated to be abandoned by her only surviving child. Sartre says that his student *"was hesitating between two kinds of morality.* […] *He had to choose between those two."* [113]

---

[112] Jean-Paul Sartre, *Existentialism Is a Humanism* (Yale University Press, 2007).
[113] Ibid, p.40.

A cognitivist interpretation of this case of moral doubt would explain it in terms of the student's partial degrees of belief about what he ought to do. He is uncertain about which of two statements is true: that he should go to war or that he should stay at home with his mother. Conversely, non-cognitivists can explain this case in terms of the student's ambivalence. He faces difficulty in deciding what to do, because he concurrently has reasons for going to war and for not going to war. Due to the circumstances he finds himself in, his desire to care for his mother conflicts with his desire to defend his country and avenge his brother. The experience of Sartre's student seems to capture the feature of certitude that Smith identifies. However, ambivalence in non-cognitive attitudes seems equally well equipped to make sense of his state of mind as cognitivist degrees of belief. While a sense of filial duty provides a compelling reason to stay, fraternal loyalty and patriotism provide reasons not to. Explaining the difficulty of this decision and the hesitation of the student does not require the attribution of moral beliefs. If anything, ambivalence seems to give a truer account of the student's experience. He does not feel merely uncertain, as one might about tomorrow's weather. He feels torn. Conflicted. Divided. The very language of moral dilemmas connotes ambivalence over uncertainty.

In addition to this example, a quick survey of the literature on moral uncertainty reveals that many of the cases these authors discuss could plausibly be redescribed in terms of attitudinal ambivalence. They cite uncertainty about the truth of liberal values or religious moral codes, or about the morality of abortion, trolley cases, and eating meat. [114] Yet all of these examples might alternatively be characterised as ambivalent non-cognitive attitudes. Moreover, the few cases that do seem unequivocally to comprise degrees of belief rather than ambivalence involve implausible moral psychology. In a recent paper by Will MacAskill and Toby Ord, for example, we are asked to consider the case of an agent who has a credence of 40% in some form of non-consequentialism and 60% in hedonic utilitarianism; she is uncertain of which is the true moral theory. [115] This certainly sounds like uncertainty rather than ambivalence, but surely nobody (a handful of moral philosophers aside) undertakes moral reflection in this way, considering each moral theory in turn and assigning it a credence. At the very least, non-cognitivists may be able to set

---

[114] Gracely, "On the Noncomparability of Judgments Made by Different Ethical Theories"; Lockhart, *Moral Uncertainty and Its Consequences*; Ross, "Rejecting Ethical Deflationism"; Sepielli, "'Along an Imperfectly-Lighted Path': Practical Rationality and Normative Uncertainty"; MacAskill, "Normative Uncertainty"; Harman, "The Irrelevance of Moral Uncertainty."
[115] MacAskill and Ord, "Why Maximize Expected Choice-Worthiness?"

this sort of case to one side as a problem particular to cognitivism. Some may find themselves in just such a quandary, but this is due to their mistaken understanding of the nature of moral judgements and it is no major loss to non-cognitivism if it cannot accommodate these situations. It is the true-to-life moral dilemmas such as Sartre's student that we should be concerned with explaining and in these cases ambivalence fares at least as well as uncertainty.

In order to serve as the missing link between non-cognitivism and moral uncertainty, however, ambivalence must be able to play the role of low certitude in moral judgements. There are a number of desiderata that comprise this role. First and foremost, ambivalence should be gradable. Moral certitude is not binary, but comes in varying degrees as individuals feel more or less confident of their moral judgements. Likewise, it seems clear that ambivalence comes in degrees. It is routinely quantified in the scientific study of attitudes and there a number of methods for its measurement. It also seems intuitively right that agents can be ambivalent to a greater or lesser extent: the stronger and more closely balanced their conflicting attitudes, the greater their degree of ambivalence.

Second, ambivalence should be able to vary through time, to give a moral judgement's robustness. This feature of ambivalence follows straightforwardly from its gradeability. For example, if Leontius were to reflect on the fact that death is a natural part of life, his feeling of disgust and desire not to look might diminish, thereby reducing his ambivalence. As an agent's reasons arise and fade away, and as the strength of these reasons waxes and wanes, there will be a consequent change in the degree to which that agent is ambivalent.

Next, certitude is distinct from importance, so ambivalence should be distinct from strength. These features are not entirely orthogonal, since the strength of conflicting desires partly determines ambivalence. However, ambivalence does not merely track strength and they are distinct variables. One may feel more ambivalent about lying out of politeness than about murder, because the competing reasons for and against the former are more closely balanced, even though the latter elicits stronger reasons overall. In fact, an increase in the strength of reasons not to murder would make one less ambivalent, not more. Likewise, certitude and importance are distinct, but not completely orthogonal. Considerations that would increase importance in the judgement that it would be wrong

to lie, such as having taken an oath to tell the truth or speaking to a friend who highly values honesty, also seem to confer additional certitude in this judgement. Suppose, however, that one were to learn of additional considerations both against and in favour of lying. Then, we might expect the matter to be considered one of greater importance, but reduced certitude. Or, as per the proposed analysis, one of stronger non-cognitive attitude and greater degree of ambivalence. Thus, although ambivalence and strength are not completely orthogonal, the relationship between them appears to be just what we should expect for an account of certitude and importance in moral judgements.

Smith describes how the strength of a rational agent's motivation covaries with the certitude of their moral judgements. Ambivalence should therefore bear this relation to motivation in order to play the role of certitude. And indeed, it seems to. If Leontius had a desire to look at a beautiful sunset, equal in strength to his desire to look at the corpses, and no conflicting desire not to, we would expect him to be more motivated to look at the sunset than the corpses. Moreover, there is strong scientific evidence to support this relationship. Numerous studies have found that a higher degree of ambivalence leads to a weaker attitude-behaviour relationship, with attitudes less predictive of behaviour. Ambivalence weakens the link from attitudes to intentions and from intentions to behaviour. This effect has been observed in a wide range of behaviours, including diet, [116] smoking, [117] sex safe behaviour, [118] and recreational drug use. [119] In summary, the greater the degree of ambivalence in an individual's attitude, the less likely they are to act on that attitude, just as should be the case for an account of certitude.

It should now be clear that ambivalence has all of the features that we would expect it to have in order to play the role of low certitude in our moral judgements. It is gradable; it

[116] Christopher J Armitage and Mark Conner, "Attitudinal Ambivalence: A Test of Three Key Hypotheses," *Personality and Social Psychology Bulletin* 26, no. 11 (2000): 1421–32; Mark Conner et al., "Moderator Effects of Attitudinal Ambivalence on Attitude–Behaviour Relationships," *European Journal of Social Psychology* 32, no. 5 (2002): 705–18; Mariëtte Berndsen and Joop Van der Pligt, "Ambivalence towards Meat," *Appetite* 42, no. 1 (2004): 71–78.
[117] Isaac M Lipkus et al., "The Relationship Between Attitudinal Ambivalence and Desire to Quit Smoking Among College Smokers," *Journal of Applied Social Psychology* 31, no. 1 (2001): 113–33.
[118] Joseph R Priester and Richard E Petty, "Extending the Bases of Subjective Attitudinal Ambivalence: Interpersonal and Intrapersonal Antecedents of Evaluative Tension.," *Journal of Personality and Social Psychology* 80, no. 1 (2001): 19.
[119] Conner et al., "Moderator Effects of Attitudinal Ambivalence on Attitude–Behaviour Relationships"; Mark Conner and Christopher J Armitage, "Attitudinal Ambivalence," in *Attitudes and Attitude Change*, ed. W Crano and R Prislin (New York: Psychology Press, 2008), 261–86.

can fluctuate through time; it is suitably distinct from importance; and it covaries with motivation. So, the presence of moral doubt need pose no problem for non-cognitivism. I do not mean to claim that ambivalence is the only possible way of understanding moral doubt. It can, after all, be interpreted as uncertainty by cognitivists. But ambivalence at least provides an account that matches the true phenomenon of moral doubt and is consistent with non-cognitivism. I shall now consider some potential objections to the claim that ambivalence can play the role of moral doubt in non-cognitivist accounts of moral judgements.

# 5    Objections

## 5.1 Ambivalence *and* Uncertainty

One may accept that there is ambivalence in moral judgements, but claim that this is in addition to, rather than instead of, moral certitude. The problem facing non-cognitivism arose because non-cognitive attitudes have fewer structural dimensions than moral beliefs. If we introduce ambivalence, this gives a richer structure to non-cognitive attitudes. But this move also produces an equivalent increase in the structural complexity of moral beliefs, since they come with their own degree of ambivalence. [120] Because ambivalence is not unique to non-cognitive attitudes, its introduction does not close the gap between them and moral beliefs. Moral beliefs can vary in certitude *and* ambivalence. Therefore, so should moral judgements.

While it is true that there may be ambivalence over and above uncertainty in beliefs, this does not mean that moral judgements must also have both. The crux of the problem was not that non-cognitive attitudes have less structure than beliefs, but that they have less structure than moral judgements seem to. In order for this objection to have force it must be shown that we should also separate uncertainty and ambivalence in moral judgements. But it is a necessary feature of non-cognitivism that there is no uncertainty per se, because moral judgements aren't beliefs. Moral judgements, being non-cognitive attitudes, are

---

[120] See, for example: David C Makinson, "The Paradox of the Preface," *Analysis* 25, no. 6 (1965): 205–7; Peter Klein, "The Virtues of Inconsistency," *The Monist* 68, no. 1 (1985): 105–35.

fundamentally different from beliefs. Why should we expect them to have the same structure? What matters is that non-cognitive attitudes can be shown to have the structure as actual moral judgements. And, as we have seen, they can. There is nothing missing from the non-cognitivist's explanation of realistic moral judgements, such as those regarding the dilemma facing Sartre's student.

## 5.2 Uncertainty About Desires

It may be objected that ambivalence is unnecessary for accommodating certitude, because an agent may simply be uncertain about their desires. Bradley and Drechsler suggest that the desirability of an outcome in a decision problem may depend not only on features of that outcome itself, but also on those of the decision maker, such as their tastes, preferences, or needs. [121] An agent may be just as uncertain about their own desires as they are about the features of an outcome itself. Similarly, Jerzak has suggested that agents may be uncertain or mistaken about their own desires. [122] Surely it is this uncertainty about our own desires that should stand in for moral doubt for non-cognitivists.

There are two reasons for non-cognitivists not to analyse moral doubt in terms of an agent's beliefs about their desires. Firstly, there seem to be examples in which agents may experience moral doubt despite being certain in their beliefs about their desires. Sartre's student is just such an example: he believes with certainty that he desires to stay with his mother and he believes with certainty that he desires to join the French resistance, but he feels moral doubt nonetheless, because of the conflict between these desires.

Secondly, an important idea underlying non-cognitivism is the Humean theory of motivation, according to which motivation stems from a desire for some end and beliefs about the how to realise it. Analysing moral doubt in terms of beliefs about desires would suggest that we act on the basis of what we believe our desires to be, rather than on the desires themselves. This is in tension with the Humeanism that underpinned non-cognitivism in the first place, so hardly provides a suitable account of moral uncertainty

---

[121] Bradley and Drechsler, "Types of Uncertainty."
[122] Ethan Jerzak, "Two Ways to Want?," *The Journal of Philosophy* 116, no. 2 (2019): 65–98.

for non-cognitivists. A truly non-cognitivist theory of meta-ethics needs an account of certitude without recourse to beliefs. Ambivalence remains the only plausible candidate.

## 5.3 A Property of What: Desires or Agents?

One potentially troubling feature of the definition of ambivalence offered above is that it describes a property of an agent, rather than an attitude. Ambivalence is the state of having attitudes that give rise to reasons for mutually exclusive options, but moral certitude is a feature of an attitude itself. This is unlike the cognitivist conception of moral uncertainty in terms of degrees of belief, since this is a property of the belief itself.

However, note that in terms of both belief and non-cognitive attitudes we talk about properties of the attitude itself *and* of the agent who holds it. With beliefs, we have a property of the attitude—degree of belief—and a property of the agent—uncertainty. An agent is uncertain concerning $p$ if they hold a partial belief in $p$. In other words, if their credence in $p$ is <1. Likewise, with non-cognitive attitudes we have a property of the attitude—conflict with other attitudes—and a property of an agent—ambivalence. An agent is ambivalent if they hold attitudes that give rise to reasons for mutually exclusive options. This also coheres with the feature of moral judgements that we are trying to understand: there is a property of the judgement—certitude—and a property of the agent who makes it—moral doubt. Thus, the fact that the discussion so far has focused on the property of the agent, rather than the property of the attitude should be no cause for concern. It is just like talking about uncertainty, rather than partial degrees of belief, or like talking about moral doubt, rather than the certitude of moral judgements.

However, a further problem may appear to linger. While the degree of a belief is an intrinsic property of that belief, the consistency of a desire with other desires is a relational property. It concerns the way in which one desire relates to the others that an agent holds. Another disanalogy, another possible weakness in the proposal. However, a disanalogy between ambivalence and uncertainty is not necessarily problematic. Indeed, in so far as this paper is attempting to propose an alternative understanding of moral doubt, it would be entirely uninteresting if there were no differences between the cognitivist and non-cognitivist

conceptions of the phenomenon. As in response to the first objection considered, the question is not whether moral ambivalence is exactly like cognitivist moral uncertainty, but rather if it adequately matches the phenomenology of moral doubt. It would be question-begging for a cognitivist to claim, without further argument, that moral doubt must be explained in terms of an intrinsic property of moral judgements, like degrees of belief.

The question, then, is that of whether we should think of moral certitude as a feature of moral judgements' relations to one another, or as an intrinsic property of each separate judgement. The examples considered throughout this paper lend some support to the view that moral doubt takes the form of conflict between different judgements, rather than an intrinsic property of each, but this support falls short of a knockdown argument. It is worth noting, however, that it only seems to be uncertainty that comes with a degree of ambivalence that poses a challenge when deciding how to act. If we are uncertain about some matter of fact that will determine the outcome of a choice between two options $A$ and $B$, but option $A$ is better than option $B$ in all possible states of the world, then this uncertainty is of little practical import. [123] Uncertainty is relevant to our practical deliberation when the different partial beliefs provide reasons for mutually exclusive courses of action. For example, If I am uncertain about whether or not it will rain, I may be unsure of whether or not to take an umbrella. Suppose that I would prefer to take my umbrella if it does rain and not if it doesn't. I believe, to degree $x$, that it will rain. This provides me with some reason to take my umbrella. However, I also believe, to degree $y$, that it will not rain. This provides me with some reason not to take my umbrella. Of course, I cannot both take my umbrella and not take it, so these conflicting reasons, which depend on my partial beliefs, make me ambivalent. If I preferred to take my umbrella regardless of whether or not it rained, my uncertainty about the rain would be irrelevant to the decision of whether or not to take it, since it would be obvious that I should. Note that the ambivalence here is not intrinsic to the belief, but is a relational property of my partial beliefs in light of the reasons for action to which they give rise. If it is this relational property that explains the practical significance of uncertainty, then there seems little

---

[123] This assumes that the states are probabilistically independent of the options. If such independence does not hold, however, then the significance of the uncertainty also depends on the conflicting reasons that would arise from the effect of the choice on the probability of the states, which is another type of ambivalence.

reason to think that a relational property is ill-equipped to make sense of the variations in moral certitude.

# 6     Conclusion

In this paper, I have presented one way of understanding the phenomenon of moral doubt that is consistent with non-cognitivism. On this view, moral doubt is not uncertainty in a person's beliefs, but ambivalence in their non-cognitive attitudes. That is, the presence of conflicting reasons that arise from these attitudes and support mutually exclusive options. The existence of ambivalence controverts the claim that non-cognitive attitudes have only two structural features. This undermines a central premise in the argument for the view that moral judgements cannot be non-cognitive attitudes due to their structural differences. The principal upshot is that one can consistently be a non-cognitivist and recognise variation in the certitude of moral judgements.

However, the differences between ambivalence and uncertainty also suggest a different understanding of, and practical response to, the phenomenon of moral doubt. As noted at the outset, there have been a number of attempts to answer the question of how to act under moral uncertainty that presuppose moral realism. But these do not seem applicable on the non-cognitivist interpretation of moral doubt as attitudinal ambivalence. One such view, defended by Elizabeth Harman and Brian Weatherson, is that moral uncertainty is irrelevant to how one should act. [124] They claim that there is an objectively true moral theory that determines what is right regardless of an agent's moral beliefs. You ought to do what you ought to do, irrespective of whether you are certain about it or not. This view crucially depends on the existence of objective moral facts, so is simply incompatible with non-cognitivism.

An alternative approach, explored in Chapter 2, is to treat moral uncertainty as equivalent to empirical uncertainty, using an analogue of expected utility theory designed to

---

[124] Harman, "The Irrelevance of Moral Uncertainty"; Weatherson, "Running Risks Morally"; Weatherson, *Normative Externalism.*

accommodate moral uncertainty. This is proposed as a norm of rationality, so does not rely on objective moral facts in a way that is inconsistent with non-cognitivism. However, the arguments that justify the basic principles of expected utility theory do not automatically apply to choices under ambivalence in the way that they are taken to apply to choices under a cognitivist conception of uncertainty. It is certainly not immediately clear, or instance, that ambivalence should, or even could, be represented like subjective degrees of belief that obey the probability axioms. Furthermore, the claim that moral doubt is not uncertainty at all, but ambivalence instead, seems to provide good reason to treat it differently from empirical uncertainty, since these are fundamentally different states.

There is, however, at least one feature of this approach that can be carried over into an account of choice under ambivalence. Some argue that we ought to deal with moral uncertainty in the same way as we deal with empirical uncertainty. Analogously, we may want to deal with ambivalence in moral judgements in much the same way as we handle conflict in our non-moral reasons. How rational agents ought to deal with ambivalent attitudes, moral or otherwise, is taken up in the Chapter 4. This is a challenge that must be met by cognitivists and non-cognitivists alike.

# CHAPTER 4

# The Balance and Weight of Reasons

## 1    Introduction

A crucial part of practical deliberation involves weighing up our reasons for alternative courses of action. An agent evaluates the considerations that stand in favour of their different options and comes to some judgement about which is best, overall. In order to make a choice, one must settle where the *balance of reasons* lies. That is to say, whether or not, and the degree to which, our reasons favour one option over another. But this is not the only dimension of our reasons that influences preference and choice. An underappreciated aspect of practical deliberation is the overall *weight of reasons* that are at play in a decision. That is to say, the total amount that is at stake in a choice, or the gravity of the reasons that contribute to an agent's evaluation of their options. When new reasons come to light, two things may change: the degree to which one option is deemed preferable to the other, and the overall gravity of the reasons that are at play.

This paper presents a philosophical exploration of this oft-ignored dimension of practical reason and the role it has to play in our understanding of preference and choice. In §2 I will explicate the distinction between the balance and weight of reasons and argue that the greater the weight of reasons underlying an evaluation, the more stable that evaluation will be when new reasons come to light. In §3 I will demonstrate how this account of the weight reasons can fill some explanatory gaps in our understanding of "hard choices"—those in which one option seems better than another in some respects, the other seems better than the first in some respects, but neither seems better overall—and the analysis of these choices in terms of options that are on a par. In particular, the weight of reasons plays a crucial role in the way in which hardness can vary by degree, while also explaining what it is about these choices that makes them genuinely hard.

First, a note of clarification on my use of the terms "reason" and "preference". When I talk about reasons, I mean normative reasons. That is to say, those considerations that

94

stand in favour of, or lend justificatory support to, the actions or attitudes of an agent. When I talk about preferences, I mean all things considered comparative evaluations of a variety of objects including options, outcomes, states of the world, or anything else that one can judge as more or less desirable. [125] The "all things" here includes all of the normative reasons that a rational agent recognises as such. On this view reasons are, in some sense, prior to preferences and contribute to their formation. As will become clear in what follows, I think of the relation between preferences and reasons as similar to the relation between beliefs and evidence; our preferences are based on, and sensitive to, our reasons in much the same way that our beliefs are based on, and sensitive to, our evidence. However, the discussion of reasons that follows does not rely on any particular view about the source of reasons themselves, nor does it assume a position on the debate between internalism and externalism about normative reasons. In particular, although I see preferences as being responsive to reasons, those reasons in turn may depend on some conative attitude of the agent in question. Whatever the source of our normative reasons, the preferences of rational agents should be responsive to them and it is the way in which this is so that is my target.

## 2      The Weight of Reasons

### 2.1     The Distinction Between Balance and Weight

Conventional accounts of preference characterise agents as either preferring one option *A* to another *B*, preferring *B* to *A*, or remaining indifferent between the two. However, this account misses some important features of the way in which agents may evaluate alternatives in a choice. Take indifference: one may be indifferent between *A* and *B* because there are no good reasons for either. For example, I have no reason to prefer the state of affairs in which the temperature in Winnipeg tomorrow is 25°C over the state of affairs in which the temperature in Winnipeg tomorrow is 27°C. I won't be anywhere near Winnipeg tomorrow and simply don't care about the weather there. With no reason to

---

[125] For a defence of this interpretation of preferences, see: Daniel Hausman, *Preference, Value, Choice, and Welfare* (Cambridge: Cambridge University Press, 2011); and Richard Bradley, *Decision Theory With a Human Face* (Cambridge: Cambridge University Press, 2017). For a critique, see: Johanna Thoma, "Judgementalism About Normative Decision Theory," *Synthese*, accessed May 14, 2021, https://doi.org/10.1007/s11229-019-02487-0.

prefer one temperature to the other, I am indifferent. Alternatively, one may be indifferent between *A* and *B* because there are good reasons of equal strength in favour of both options. For example, Buridan's ass finds itself exactly half way between two equally large and delicious piles of hay; with equally strong reasons in favour of each pile of hay, the ass is indifferent between the two. Lastly, one may be indifferent between *A* and *B* because there are good reasons of equal strength *against* both options. For example, when sailing home from the Trojan War through the Strait of Messina, Odysseus has to plot a course either within reach of Scylla, a sea monster with six heads full of sharks' teeth, or Charybdis, a whirlpool that swallows into the deep whole ships and their crews. With equally strong reasons against each option, Odysseus is indifferent between the two, stuck between a rock and a hard place.

These are all instances of indifference, but there is an important dissimilarity between the first case—concerning my indifference about the weather in Winnipeg—and the other two. Whereas my indifference regarding the temperature in Winnipeg stems from my complete lack of interest in the alternative possibilities, Buridan's ass and Odysseus have a considerable amount at stake and compelling, but conflicting reasons pulling them in opposing directions. This difference is not reflected in the balance of reasons, because, in terms of the overall comparative evaluation, they are all simply classified as indifferent. Rather, this difference is captured by the fact that the latter two are ambivalent, with strong reasons for mutually exclusive options.

A common metaphor for ambivalence sees competing reasons balancing on either side of a set of weighing scales. Reasons for one option accrue on one side, reasons for an alternative on the opposite side, and the weightier sum of reasons tips the balance in favour of the option to be pursued. [126] Different comparisons might differ from one another in terms of the balance of reasons, which is represented by the relative heights of the two sides of the scales. The more that reasons for one option outweigh those for an alternative, the further down that option is pushed. However, they may also differ in the total amount

---

[126] Errol Lord and Barry Maguire, "An Opinionated Guide to the Weight of Reasons," in *Weighing Reasons* (Oxford: Oxford University Press, 2016), 3–24, https://doi.org/10.1093/acprof:oso/9780199315192.003.0001; Garrett Cullity, "Weighing Reasons," in *The Oxford Handbook of Reasons and Normativity*, ed. Daniel Star (Oxford: Oxford University Press, 2018), 423–42, https://doi.org/10.1093/oxfordhb/9780199657889.013.19; Barry Maguire and Justin Snedegar, "Normative Metaphysics for Accountants," *Philosophical Studies* 178, no. 2 (February 1, 2021): 363–84, https://doi.org/10.1007/s11098-020-01435-w.

of weight on the scales. When new considerations come to light that are relevant to a choice, they may shift the balance in favour of one option or the other, but either way the overall weight of reasons will have increased. Although my attitude towards the weather in Winnipeg and the attitude of Buridan's ass towards the piles of hay are the same in terms of the balance of reasons—we are both indifferent—they differ in terms of the overall weight of reasons.

It should be noted that the weighing scales metaphor is only a crude model of practical reason and is not intended to capture every aspect of rational deliberation. It should not be taken to imply that our reasons always simply add their weights together, like inert pieces of cast iron, nor that they simply provide unqualified support for one option or an alternative. For example, Justin Snedegar has argued that we should distinguish between reasons *against* one option and reasons *for* the alternatives,[127] while Chris Tucker has argued that a dual scale model of reasons is needed to adequately represent ethical theory. [128] More complex setups such as these might be required for some applications of the scale metaphor. Nevertheless, the basic version of this model can be fruitful for representing certain aspects of the way in which we consider and compare our practical reasons. In particular, it allows us to get an intuitive grasp of the distinction between the balance and weight of reasons, and will be a helpful tool for exploring some of the explanatory and normative implications of this distinction.

## 2.2    The Weight of Evidence and Stability of Credences

Recognising the role played by the weight of reasons, and its distinction from balance, has a number of important implications. [129] The main point that I aim to demonstrate here is that the weight of our reasons affects the stability of our preferences. In particular, the greater the total weight of reasons on the scales, the less that a new reason will be able to change the balance. This idea mirrors a closely related point in epistemology, about the

---

[127] Justin Snedegar, "Reasons For and Reasons Against," *Philosophical Studies* 175, no. 3 (2018): 725–43, https://doi.org/10.1007/s11098-017-0889-2.
[128] Chris Tucker, "The Dual Scale Model of Weighing Reasons," *Nous*, 2021, https://doi.org/10.1111/nous.12361.
[129] For example, the weight of reasons partly determines the degree of ambivalence, as it was presented in Chapter 3.

balance and weight of evidence. [130] Consider the following pair of cases, suggested by Karl Popper to illustrate his famous "*paradox of ideal evidence*": [131]

**Coin toss A**

*I hand you a coin and tell you that it is fair. A quick look at the coin suggests no reason to think otherwise. What is your credence that the 6$^{th}$ toss of the coin will land heads? You might say ½.*

**Coin toss B**

*I hand you coin and tell you that it is fair. Again, a quick look at the coin suggests no reason to think otherwise. This time, however, you play with the coin for a while, flipping it repeatedly and keeping track of the results. Over 100 coin tosses, 50 land heads and 50 land tails. What is your credence that the 106$^{th}$ toss of the coin will land heads? You might say ½.*

In some sense, you have said the same thing in response to both of these questions. In both cases, you have a credence of ½ in the proposition that the coin will land heads on some future toss. However, your credal state might well be thought to be quite different in the two cases. That difference lies in the weight of evidence that underlies your credences. In the first case, the only evidence you have is the appearance of the coin, my claim that it is fair, and your general knowledge of how coins tend to behave. In the second case, however, you have all of this evidence, plus that gained during the additional 100 coin tosses. This provides more evidence overall, but the balance of the evidence does not change. So, the balance of evidence is the same for both, but the weight of evidence is different. In the words of C. S Peirce:

---

[130] See, for example, Charles Sanders Peirce, "The Probability of Induction," in *The Essential Peirce: Selected Philosophical Writings. Volume 1 (1867–1893)*, ed. Nathan Houser and Christian Kloesel (Bloomington: Indiana University Press, 1992), 155–69; John Maynard Keynes, *A Treatise on Probability* (London: Macmillan, 1921); Rudolf Carnap, *Logical Foundations of Probability* (Chicago: University of Chicago Press, 1950); Jeffrey, *The Logic of Decision*; Brian Skyrms, "Resiliency, Propensities, and Causal Necessity," *The Journal of Philosophy* 74, no. 11 (1977): 704–13; Isaac Levi, *The Enterprise of Knowledge* (Cambridge: MIT Press, 1980); Isaac Levi, "The Weight of Argument," in *Fundamental Uncertainty: Rationality and Plausible Reasoning*, ed. S. M. D. Brandolini and R. Scazziere (Basingstoke: Palgrave Macmillan, 2011), 39–58; Peter Gärdenfors and Nils-Eric Sahlin, "Unreliable Probabilities, Risk Taking, and Decision Making," *Synthese* 53, no. 3 (1982): 361–86; Brian Weatherson, "Keynes, Uncertainty and Interest Rates," *Cambridge Journal of Economics* 26, no. 1 (2002): 47–62; Joyce, *The Foundations of Causal Decision Theory*.
[131] Karl Popper, *The Logic of Scientific Discovery (1959)* (London: Routledge Classics, 2002).

*"[…] we ought not to have the same feeling of belief in reference to all events of which the chance is even. In short, to express the proper state of our belief, not one number but two are requisite, the first depending on the Inferred probability, the second on the amount of knowledge on which that probability is based"* [132]

To see what difference the weight of evidence makes in practice, consider what happens to your credence in the 6th or 106th coin toss coming up heads as you observe tosses 1 to 5 or 101 to 105. Suppose that both sequences produce 5 consecutive heads. In the first case, your credence in the hypothesis that the 6th toss will land heads is going to increase quite substantially. The fact that all observed coin tosses have landed the same way up suggests that the coin is biased towards heads, which would increase the probability of heads on the next toss. In the second case, however, when you've already tossed the coin many times and seen an even distribution of heads and tails, a sequence of 5 heads in a row isn't going to change your credence all that much. In fact, over a long run of coin tosses, we would expect to find just this sort of sequence of back-to-back heads every so often. Because the evidence in the first case is weightier, the associated credence is more stable. That is to say, a given piece of new evidence elicits a smaller change in one's credence.

Note that in this pair of cases, the balance of evidence for *some* propositions might differ. For example, in the first case, as I have suggested, the balance of evidence suggests that the coin is likely to be biased, but in the second case it does not. Nonetheless, the cases allow us to see that there are some propositions for which the balance of evidence is the same, but the weight is different, and this will manifest as greater stability in the credence produced by the weightier body of evidence.

## 2.3    The Weight of Reasons and Stability of Preferences

What I aim to demonstrate here is that something similar should be thought to happen when it is the weight of our practical reasons rather than our evidence—or reason for belief—that differs. That is to say, when making comparative evaluations of alternatives in a choice, the greater the weight of reasons being considered, the less that a new reason will

---

[132] Peirce, "The Probability of Induction.", p.160.

be able to tip the scales in either direction. [133] To make this point clear, consider an agent evaluating two pairs of lotteries in which the outcome depends on whether a red ball or a yellow ball is pulled out of an urn, with the possible outcomes specified in Table 1.

**Table 7.** Two pairs of lotteries

|  | Red | Yellow |
|---|---|---|
| Lottery I | £100 | £0 |
| Lottery II | £50 | £50 |
| Lottery III | £1000 | £0 |
| Lottery IV | £800 | £800 |

Suppose that this agent's preferences can be represented by a cardinal utility function that, for simplicity, increases linearly with money. In other words, the only reasons underlying their preferences concern money, and every £1 contributes equally strong reasons to every other. The degree of preference for £100 over £50 is greater than that for £1000 over £800. This difference in strength of preference will manifest in this agent's choices between these lotteries with different probabilities of pulling out red and yellow balls: if offered a choice between Lottery I and Lottery II, the agent would be indifferent when the probability of a red ball $p$(Red)=0.5, prefer Lottery I when $p$(Red)>0.5, and prefer Lottery II when $p$(Red)<0.5. However, if offered a choice between Lottery III and Lottery IV, the agent would be indifferent when $p$(heads)=0.8, prefer Lottery III when $p$(Red)>0.5, and prefer Lottery IV when $p$(Red)<0.5. They will accept a lottery with a lower probability of a good outcome in the first choice, because the degree to which the good outcome in the lottery is preferred to the guaranteed alternative is greater in the first choice than in the second.

However, their preferences over the possible outcomes in Lottery III and Lottery IV are, in some sense, more stable than their preferences over the possible outcomes in Lottery I and Lottery II. Even though the strength of preference for £100 over £50 is greater than that for £1000 over £800, the outcomes will need to be changed by greater amount in

---

[133] For a discussion of how reasons combine, see: Jonathan Dancy, *Ethics Without Principles* (Oxford: Oxford University Press, 2004); Campbell Brown, "The Composition of Reasons," *Synthese* 191, no. 5 (2014): 779–800, https://doi.org/10.1007/s11229-013-0299-8.

order to elicit a change in the latter preference than in the former, because the overall amount at stake is larger. That is to say, the amount by which the outcome of picking a red ball after choosing Lottery II would have to be improved in order to make it preferable to the outcome of picking a red ball after choosing Lottery I, is greater than the amount by which the outcome of picking a red ball after choosing Lottery III would have to be improved in order to make it preferable to the outcome of picking a red ball after choosing Lottery I. That amount is $>£50$ in the first pair of lotteries and $>£200$ in the second pair of lotteries. Thus, a fixed change to the possible outcomes of the choice between Lottery I and II will elicit a smaller change in the agent's preferences than the same change to the equivalent outcomes in the choice between Lottery III and IV. This will also manifest in smaller changes to the probability of picking a red ball at which the agent will be indifferent between the lotteries, for a given change in the possible outcomes.

The effect of the amount at stake on the stability of one's preferences can also be seen in richer, more realistic decisions than these simple toy examples. Consider another pair of choices.

> **Moving house**
>
> *I recently moved house and had to decide where to live. I narrowed down the options to two: remain in London, where I have always lived, or move to Otley, West Yorkshire. In London I have family and friends, and access to the outstanding arts and culture of a global metropolis. Otley, on the other hand, presents the opportunity to live within touching distance of the Yorkshire Dales and the excitement of getting to know somewhere new. On balance, I prefer Otley.*

> **Dessert**
>
> *I must choose whether to have vanilla or chocolate ice cream after dinner this evening. The only reason I have for either is the amount of gustatory pleasure I would derive from each. Although I like ice cream, there simply aren't very weighty reasons on the scales here. On balance, I prefer vanilla.*

In virtue of the weightier reasons involved, my preference in the choice of where to live will be more stable than in the choice of which dessert to have. That is, one of the options would have to be improved by a greater degree for my preference ordering to change. For

example, although I prefer vanilla to chocolate, I would prefer chocolate plus £100 to vanilla and no monetary bonus. This is not the case in the choice between London and Otley; adding £100 to the option of living in London will not reverse my preference for Otley. In fact, to change that preference ordering would require a far more substantial change to the reasons for the two options. And this seems to be exactly how things should be! Evaluating these alternatives involved a comparison between weighty reasons. It felt like an important, high stakes decision, with substantial considerations in favour of each option. Therefore, it would seem absurd for one's preferences regarding this important life choice to be so easily reversed by a comparatively small monetary sweetener. Thus, the analogy with the weight of evidence holds: when considering a great deal of evidence to reach a certain credence, that credence is more stable in the face of new evidence, and when there are strong reasons at stake when making comparative evaluation between choice options, this evaluation will be more stable in the face of new reasons.

## 2.4    Potential Objections

There are a number of ways that one might be tempted to explain the difference between these preferences without reference to the weight of reasons. First, one may claim that money simply doesn't feature as a reason for me in the decision of where to live, so financial incentives can't tip the balance. It may be tempting to think that important personal matters like this should depend on considerations more profound and important than a bit of extra cash. This line of thought, however, does not hold up to much scrutiny. Even if we were to stipulate in the case that a significant reason to prefer Otley over London was the far lower cost of living, this would not force us to accept that adding £100 to the option of living in London would reverse my preference for Otley. Financial concerns are often a pivotal factor for even highly personal choices, but this does not mean that such choices can always be swayed by any given monetary improvement to one of the options.

Alternatively, one might think that the qualitative difference between the options makes it difficult to assess the strength of additional reasons, so the overall change needs to be bigger. Because London and Otley are qualitatively very different, it is hard to compare the reasons for living in each with enough precision for something like an additional £100 to make a difference, but chocolate and vanilla ice cream are sufficiently similar for the choice

to be sensitive to this additional reason. While it is true that qualitative difference makes evaluative comparisons less precise, as I will discuss in more detail in §3, this does not fully account for the asymmetry between these cases. Suppose instead I were to choose between moving to Otley and moving to Ilkley, another market town between the city of Leeds and the Yorkshire Dales. Life in these two towns is fairly similar, but I have a preference for Otley over Ilkley, because, among other things, it has better pubs. £100 still wouldn't be enough to swing this decision, even though the options are qualitatively so similar. However, if I were instead choosing between vanilla ice cream and a cigarette after dinner, I might clearly prefer the ice cream even though they are qualitatively very different, but this preference *could* be reversed by adding a £100 bonus to one of the options. In other words, I would prefer the ice cream to the cigarette, but would prefer the cigarette plus £100 to the ice cream. So, even when the options are more qualitatively similar, the effect of additional reasons is smaller if the overall weight of reasons is greater.

Lastly, it may appear as though the difference between these cases can simply be captured in the balance of reasons, with no need to complicate matters with the distinction between balance and weight. Perhaps the degree to which my reasons favour Otley over London is greater than that for vanilla over chocolate. This is not right either. Suppose instead we were to compare my preference for Otley over London to my preference for vanilla over banana ice cream. Although I in fact prefer Otley to London, the reasons for each are fairly evenly balanced and it feels like something of a close call. My preference for vanilla over banana, on the other hand, is clear cut, as I don't enjoy banana ice cream anywhere near as much as vanilla. Even then, while my preference for Otley over London could not be switched by £100, my preference for vanilla over banana could. £100 in the context of the weak reasons that settle ice cream choice has far more sway than £100 in a decision as important as where to live. Even when the balance of reasons is more uneven in one comparison than another, the comparison with the greater overall weight of reasons will be less sensitive to additional reasons.

The difficulty with applying this idea to preferences is that standard accounts of preference don't provide rich enough information to quantify the degree of preference for one option over another, nor the absolute value of an option. The only information taken to be captured by preferences is an ordering of alternatives and, at most, information about the ratios of value difference between different options. For example, we can say that the

difference between *A* and *B* is twice the difference between *C* and *D*, but we can't say that *A* is twice as good as *B*, nor that *A* has an absolute value of *x* and *B* has an absolute value of *y*. This is what is captured by the familiar idea in decision theory that the cardinal utility functions used to represent preference orderings are unique only up to positive affine transformations. Talking in terms of the balance and weight of reasons, on the other hand, gives an intuitive picture of the idea. We have, at the very least, an intuitive grasp of what it means to say that the balance of reasons in the choice of dessert is similar to that in the choice of where to live, but that the overall weight of reasons is far greater in the latter choice.

How one defines a precise measure on the balance and weight of reasons will depend on the specific view one adopts about the nature of reasons themselves. Whether you think that a reason to φ is evidence that one ought to φ, [134] a premise in good reasoning that leads one to φ, [135] an explanation of why it is good or right to φ, [136], a primitive notion [137], or something else entirely, one will need to provide a different account of how to define and measure the balance and weight of reasons. However, the motivation for distinguishing these two dimensions of practical reason do not depend on any particular conception of reasons. In light of this, the ability to capture balance and weight as distinct variables may be considered a desideratum of any full account of reasons. Whether any particular account is better able to accommodate this distinction than others remains an open question for future work. However one defines a measure on the weight of reasons, it is my contention that this feature of practical deliberation has important explanatory implications for a

[134] Stephen Kearns and Daniel Star, "Reasons: Explanations or Evidence?," *Ethics* 119, no. 1 (October 2008): 31–56, https://doi.org/10.1086/592587; Stephen Kearns and Daniel Star, "Reasons as Evidence," in *Reasons as Evidence*, ed. Russ Shafer-Landau, vol. 4 (Oxford: Oxford University Press, 2009), 215–42; Matthew Silverstein, "Reducing Reasons," *Journal of Ethics and Social Philosophy* 10, no. 1 (June 7, 2017): 1–23, https://doi.org/10.26556/jesp.v10i1.95; Judith Jarvis Thomson, *Normativity* (Chicago: Open Court, 2008); Daniel Whiting, "Right in Some Respects: Reasons as Evidence," *Philosophical Studies* 175, no. 9 (2018): 2191–2208, https://doi.org/10.1007/s11098-017-0954-x.

[135] P Hieronymi, "The Wrong Kind of Reason," *Journal of Philosophy* 102, no. 9 (2005): 437–57, https://doi.org/10.5840/jphil2005102933; Conor McHugh and Jonathan Way, "Fittingness First," *Ethics* 126, no. 3 (April 1, 2016): 575–606, https://doi.org/10.1086/684712; Kieran Setiya, "What Is a Reason to Act?," *Philosophical Studies* 167, no. 2 (2014): 221–35, https://doi.org/10.1007/s11098-012-0086-2; Jonathan Way, "Reasons as Premises of Good Reasoning," *Pacific Philosophical Quarterly* 98, no. 2 (2017): 251–70, https://doi.org/10.1111/papq.12135.

[136] Stephen Toulmin, *En Examination of the Place of Reason in Ethics* (Cambridge: Cambridge University Press, 1950); Stephen Finlay, *Confusion of Tongues: A Theory of Normative Language* (Oxford: Oxford University Press, 2014); Joseph Raz, *Engaging Reasons* (Oxford: Oxford University Press, 1999); John Hyman, *Action, Knowledge, and Will* (Oxford: Oxford University Press, 2015); John Broome, *Rationality Through Reasoning* (Oxford: Wiley-Blackwell, 2013).

[137] Thomas Scanlon, *What We Owe to Each Other* (Cambridge: Harvard University Press, 1998); Derek Parfit, "Rationality and Reasons," in *Exploring Practical Philosophy*, ed. Dan Egonsson et al. (Aldershot: Ashgate, 2001), 17–41.

particular class of decisions known as "hard choices". There are a number of well-known proposals for how such choices should be analysed and what implications they have for practical rationality. In what follows, I will on focus one such proposal: that hard choices are those between options that are "on a par". I will suggest that this account leaves open some important gaps in our understanding of the kind of decision at stake, but that these gaps can be filled once we recognise the role of the weight of reasons. In combination, the ideas of parity and the weight of reasons provide the conceptual tools to explain hard choices, and lay the foundation for how agents can deliberate and act in this context.

# 3    Hard Choices

## 3.1    Incomparability and Parity

There are many ways in which a decision may be difficult. For example, it might be cognitively demanding, emotional painful, or require significant willpower. However, there is a particular class of decisions that have become known as hard choices, following Isaac Levi's 1986 book of the same name.[138] These are choices in which one option seems better than another in some ways, the other seems better than the first in some ways, but neither seems better than the other, all things considered.[139] The different "ways" in which the options are better than one another concern different types of value that they exhibit, or different kinds of reason that support them. [140] This presents a puzzle for our understanding of the ways in alternatives can relate to one another.

When making quantitative comparisons between two objects, say *A* and *B*, it is natural to think that either *A* is greater than *B*, *B* is greater than *A*, or else they are equal. This most obviously applies for simple physical quantities: either I am taller than you, you are taller than me, or we must be the same height. The same is also true of more complex, abstract quantities, such as the price of a good, the mortality rate of a disease, or the number of words in a language. When evaluating options in a decision, however, things are not always

---

[138] Isaac Levi, *Hard Choices* (Cambridge: Cambridge University Press, 1986).

[139] For simplicity, I will stick to decisions with two options, but in reality, of course, individuals may face hard choices between any number of possible alternatives.

[140] For a related discussion of the structure of value and the structure of practical reasons, see Ruth Chang, "Parity, Imprecise Comparability and the Repugnant Conclusion," *Theoria* 82, no. 2 (2016): 182–214, https://doi.org/10.1111/theo.12096.

so simple. Sometimes it seems that neither of two alternatives is better than the other, but nor are they equally good. A classic example of this, suggested by Joseph Raz, is that of a person leaving school and trying to choose what career to pursue. [141] They have an excellent academic record and a place at a prestigious law school, but are also a talented musician and stand a good chance of making it as a professional clarinettist. So they ask themself, which is better: a successful career as a lawyer, or a successful career as a clarinettist? The legal career is better than the career as a clarinettist in terms of salary, job security, and social impact. Becoming a clarinettist, on the other hand, is better in terms of creative expression and aesthetic appeal. It might well be said that it is neither true that the law career is better than the musical career, nor that the musical career is better than the law career. In this case, however, that doesn't necessarily mean that they are considered equal. For if that were the case, then any small improvement to either would settle the choice. But we could add £100 to the hypothetical salary of the law career and it would still not seem as though this option was better than the alternative. The extra money would improve this option, but it would not break the tie, so the options must not be precisely equally good.

The conclusion Raz draws from this example is that the two are incomparable. The reason that neither is better than the other, but nor are they equally good as career options, is that they are just too different to be compared. According to this view, hard choices are those in which the options are incomparable, and two options are incomparable if and only if it is neither true that one is better than the other, nor true that they are of equal value. This definition reveals a tacit assumption that better than, worse than, and equal to are the only comparative relations that can hold between options. However, a number of philosophers have claimed that there are other ways in which options may be related. Prominent among these is Ruth Chang, who has argued that there are situations in which neither option is better or worse than the other, nor are they equally good, but they are comparable, since they are "*on a par*".[142] The idea is that options may be in the same overall neighbourhood of value, in some respect, but be so qualitatively different to one another that they are not related by any of the three conventional comparative relations. Two options that are on a par both exhibit some particular type of value, but in very different ways. They are in the

---

[141] Joseph Raz, *The Morality of Freedom* (Oxford: Clarendon Press, 1986).
[142] Chang, "Hard Choices"; Ruth Chang, "The Possibility of Parity," *Ethics* 112, no. 4 (2002): 659–88, https://doi.org/10.1086/339673; Ruth Chang, "Are Hard Choices Cases of Incomparability?," *Philosophical Issues* 22, no. 1 (2012): 106–26.

same overall neighbourhood of value, because the comparison between them does not favour either one over the other. Therefore, neither is better, nor worse, than the other. Nonetheless, they are not equally good, because, in light of their qualitative dissimilarity, there is an evaluative difference between them. So, they are comparable, despite neither being better than, worse than, nor equal to, the other. They are on a par. The idea of being in the same overall neighbourhood of value is somewhat vague, and I will suggest a more precise definition in §3.2, but this general overview should suffice for present purposes. [143]

The concept of parity can be elucidated by the very case that Raz used to illustrate incomparability. Neither the career as a lawyer, nor the career as a clarinettist is better than the other in terms of goodness as a career. And this tie cannot be broken by adding £100 to the salary of the law career, so they are not equally good. However, this does not mean that there is *no* possible tie-breaker. If we were to increase the salary of the law career in the comparison by ten times, this might well make this option decisively better than the alternative. Or, if we stipulated that the law career was guaranteed to result in redundancy and subsequent unemployment mid-career, this too might settle the matter. The reasons in favour of becoming a clarinettist are more closely balanced against those for the legal career in the original example than they are for these alternative versions of the legal career that break the tie. Clearly, then, we can say *something* about the comparison between the original options. Since they both exhibit a degree of goodness as a career, but neither is better than the other, they are in the same overall neighbourhood of value. But the significant qualitative differences between them mean that they are not exactly equal. Rather, they are on a par.

Choices between options that are on a par are hard because, although neither option is better than the other, this doesn't mean that there is nothing at stake, as would be the case if they were of exactly the same value. In contrast, when choosing between options that

---

[143] A third way of understanding Hard Choices is in terms of indeterminacy. The idea is that comparative evaluations are vague, so there are some cases in which it is neither true nor false that one option is better than the other. For arguments in favour of this view see: John Broome, "Is Incommensurability Vagueness?," in *Incommensurability, Incomparability, and Practical Reason*, ed. Ruth Chang (Cambridge: Harvard University Press, 1997); John Broome, "Incommensurable Values," in *Well-Being and Morality: Essays in Honour of James Griffin*, ed. Roger Crisp and Brad Hooker (Oxford: Clarendon Press, 2000); and Cristian Constantinescu, "Value Incomparability and Indeterminacy," *Ethical Theory and Moral Practice* 15, no. 1 (2012): 57–70. For criticism of this view, see: Ruth Chang, "The Possibility of Parity," *Ethics* 112, no. 4 (2002): 659–88; Erik Carlson, "Broome's Argument against Value Incomparability," *Utilitas* 16, no. 2 (2004): 220–24, https://doi.org/10.1017/S0953820804000548; Erik Carlson, "Vagueness, Incomparability, and the Collapsing Principle," *Ethical Theory and Moral Practice* 16, no. 3 (2013): 449–63, https://doi.org/10.1007/s10677-012-9352-9.

are precisely equally balanced, such as which one of multiple copies of the same book to pick from a shelf, neither option is better than the other, but this choice is not hard, because it doesn't really matter which copy one picks; the choice can be made in a completely arbitrary, thoughtless way. The qualitative difference between the value of alternatives that are on a par means that the choice matters, but the reasons on the table fail to settle which of the options is to be pursued. This, so the reasoning goes, is what makes "hard choices" hard.

## 3.2    Degrees of Hardness

The concept of parity appears to provide a compelling explanation of hard choices. As it stands, however, this account has a substantial weakness, in that it treats as equivalent choices that differ in important ways. For Chang, parity, and therefore hardness, seems to be binary. Choices are either hard, if the options are on a par, or not otherwise. But this ignores some important differences between various choices in which the options are on a par. Consider the example that Chang uses throughout her paper "Hard Choices": a choice between apple pie and lemon sorbet for dessert. [144] While apple pie is deliciously warm and comforting, and lemon sorbet is deliciously tart and refreshing, neither seems tastier overall. Chang shows how this is a choice between options that are on a par: neither is better than the other, and nor are they equal, but they are comparable, since they are in the same overall neighbourhood of value. In this case, we can test that they aren't precisely equal, because a dollop of whipped cream would improve the apple pie, but still might not break the tie with the lemon sorbet, because of the significant qualitative difference between them. To borrow a phrase from Caspar Hare, the choice is (quite literally) "*insensitive to mild sweetening*". [145] This analysis puts the choice of dessert on an equal footing with other hard choices, such as Raz's example of the choice between a career as a lawyer and a career as a clarinettist. But this doesn't seem right. For one thing, the choice between apple pie and lemon sorbet just isn't very hard at all. It isn't the sort of choice that people tend to agonise over or worry about very much, and certainly not to the degree that seems proportionate in the decisions at life's major junctures. We often settle choices like the one Chang considers in an arbitrary way, without any real trouble. If someone were to be stuck when picking what to order in a restaurant and left the choice to the toss of a coin, we

---

[144] Chang, "Hard Choices."
[145] Caspar Hare, "Take the Sugar," *Analysis* 70, no. 2 (2010): 237–47, https://doi.org/10.1093/analys/anp174., p.237.

would not be dismayed by their disregard for the import of the matter in the way that we might were someone to adopt that approach to choices such as whether to pursue a particular career, marry a partner, or conceive a child. Yet Chang's account of parity provides no clear way to distinguish between these choices. There seems to be a gradable dimension of hard choices that is missing from the analysis in terms of parity. Moreover, if the choice of career is a harder choice than the choice of dessert, but both choices meet the criteria for parity, then parity alone cannot explain why the former is so much harder than the latter. Given that some choices between options that are on a par are much harder than others, this account fails both to meet the need for a graded notion of hardness, as well as to provide an explanation of what it is that makes the truly hard choices so hard.
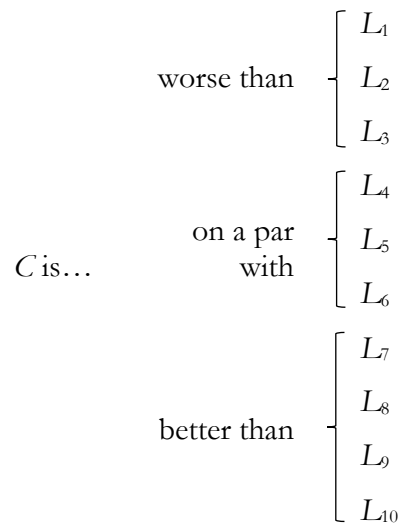
This presents a problem for an analysis of hard choices based on the parity relation, but this problem is surmountable. Once we properly understand the balance and weight of reasons, we can fill in the remaining gaps in our understanding of hard choices. In order to see this, however, it is important first to add one final detail to the concept of parity: not only can two options be on a par, but there is a range of values over which variations on one option are on a par with a fixed alternative. This point is illustrated by thinking about the variations on the legal career considered above. When compared to the career as a clarinettist, there are some versions of the legal career that are better and some that are worse, but there are multiple variations on this option that fall within a zone of parity with respect to the alternative. John Broome has suggested an illuminating way of thinking about the structure of betterness orderings in this way. [146] Although Broome was illustrating his claim that there are zones of *indeterminacy* in the betterness relation, the same point can be made about zones of *parity* instead. [147] The idea works as follows: suppose we are comparing two alternatives—a career as a lawyer and a career as a clarinettist. We hold a standard version of the career as a clarinettist, $C$, fixed. We imagine a series of $n$ variations on the legal career, $L_1$-$L_n$, and rank them from best to worst. We can then proceed through this series of variations on the legal career, comparing each in turn to the alternative, $C$. This will divide $L_1$-$L_n$ into three zones: those that $C$ is worse than, those that $C$ is better than, and those that $C$ is neither better nor worse than (see Figure 1). If this central zone contains a single item, then that item and $C$ are precisely equally good, since any change to that item for the better or worse will change the comparative relation to either better than

[146] Broome, "Is Incommensurability Vagueness?"
[147] Chang, "The Possibility of Parity," 2002.

or worse than, thereby breaking the tie. If, however, there are multiple items in this zone, then they will be neither better nor worse than the alternative, but nor will they be equally good either. Instead, they are on a par. Both the original legal career and the legal career with an extra £100 fall within this zone of parity, while the legal career with a salary that is ten times greater and the legal career that will lead to redundancy and unemployment both fall outside of this range, as one is deemed to be better than the alternative, and the other worse.

$$
C \text{ is}\ldots
\begin{cases}
\text{worse than} & \begin{bmatrix} L_1 \\ L_2 \\ L_3 \end{bmatrix} \\[2ex]
\begin{array}{c}\text{on a par} \\ \text{with}\end{array} & \begin{bmatrix} L_4 \\ L_5 \\ L_6 \end{bmatrix} \\[2ex]
\text{better than} & \begin{bmatrix} L_7 \\ L_8 \\ L_9 \\ L_{10} \end{bmatrix}
\end{cases}
$$

**Figure 1.** Comparisons between a fixed option and variations on an alternative

With these ideas in mind, we are now in a position to provide a more precise definition of parity, by differentiating it from other comparisons in which neither option is better than the other—ties—in terms of the way in which the tie might be broken. First, when neither is better than the other, but any improvement to one option relative to the other would be enough to break the tie, then they are equally good. For example, consider an individual choosing which bureau de change at which to buy some foreign currency and evaluating the alternatives solely on the basis of the exchange rate on offer. For two bureaus that offer the same exchange rate, neither is better than the other: they are tied. This tie is in fact a case of equality, since any improvement to the exchange rate offered by either would be enough to break the tie and settle the choice. Second, when neither of two options is better than another and no conceivable improvement to either could break the tie, then they are completely incomparable. For example, it doesn't make sense to say that a particular bureau de change is better or worse than a career in law, and there is no improvement to the exchange rate or salary on offer that could settle the comparison. The

objects are simply not comparable. [148] Lastly, there are those cases in which it is not the case that *any* improvement to either option would break the tie, but there is *some* conceivable improvement that would. In the comparison between musical or legal careers, it is not the case that any improvement would break the tie, since adding £100 to the salary in the legal career does not decisively make it the better option, but there are some improvements that could break the tie, such as much larger changes to the salary or career prospects. These are cases of parity, which we can now define as follows: two options are on a par with respect to one another if (i) neither option is better than the other; (ii) either option could be made better than the other by some conceivable improvements; and (iii) it is not the case that one option would be made better than the other by any improvement.

This classification of the different ways in which a comparative evaluation may result in a tie indicates that the differences between them are really just a matter of degree. Precise equality and complete incomparability are simply the extremities on a spectrum of tied pairs of options, varying in terms of the strength of tie-breaker required to settle the matter. This immediately suggests a way of making sense of the gradable nature of hard choices that I claimed was missing from Chang's account in terms of parity: the hardness of a choice between options that are on a par depends on how different things would have to be in order for the choice to be settled. That is to say, the degree to which one of the options would have to be improved in order to break the tie between them. In the final sub-section, I will demonstrate how the degree of hardness, thought of in this way, depends on two factors—the degree of qualitative difference and the weight of reasons—and how these, rather than parity alone, are crucial for an adequate explanation of hard choices.

### 3.3    Weight of Reasons, Qualitative Difference, and Hardness

The degree of hardness in a choice depends on the amount by which one of the options would have to be improved in order for one to be strictly better than the other. When any improvement to either option would be enough to break the tie, the options are equally good. When no conceivable improvement to either could break the tie, the options are completely incomparable. Between these two extremes lie all pairs of options that are on

---

[148] This sort of comparison that is not well-defined is referred to by Chang as non-comparability

a par. But what determines how much an option would have to be changed in order to break the tie between options that are on a par? This depends on two factors.

First, the size of the zone of parity within which the options lie. The wider the range of values over which variations on one option are on a par with respect to a fixed alternative, the more that option will have to be changed in order to break the tie. Options that are on a par can be compared, since they are in the same overall neighbourhood of value, but this comparison is, in some sense, imprecise.[149] The less precise the comparison, the wider the range of values over which variations on the options are on a par. This, in turn, depends on the degree of qualitative difference between the two. Why should we think that the degree of precision depends on the degree of qualitative difference between the objects being compared? This claim is simply an extension of the idea that it is the qualitative differences between alternatives that makes comparisons between them imprecise in the first place. As Derek Parfit says,

> "*When two things are qualitatively very different, these differences would often make it impossible either that one of these things is better than the other by some precise amount, or that both things are precisely equally good.*" [150]

If qualitative difference gives rise to imprecision in the comparison between items, it is natural to think that *greater* qualitative difference will result in a *greater* degree of imprecision, reflected in the size of the zone of parity.

Second, the amount by which one of the options would have to be improved in order for a tie to be broken depends on the stability of the evaluation of each alternative in question. That is to say, the degree to which one's evaluation of some object changes in response to new reasons that stand for or against it. And this, as I argued in §2, depends on the overall weight of reasons. The weightier one's reasons when evaluating alternatives, the less sensitive this evaluation will be in response to new reasons, so the stronger those new reasons will have to be in order to break a tie.

---

[149] Chang, "Parity, Imprecise Comparability and the Repugnant Conclusion"; Derek Parfit, "Can We Avoid the Repugnant Conclusion?," *Theoria* 82, no. 2 (2016): 110–27, https://doi.org/10.1111/theo.12097.
[150] Parfit, "Can We Avoid the Repugnant Conclusion?" p.113.

In summary, if you are considering one option that falls within the zone of parity with respect to another, how easy is to break this tie depends on the extent of this zone and how sensitive the evaluation is to additional reasons. How far it has to go and how easy it is to move. And these factors, in turn, depend on the degree of qualitative difference and the overall weight of reasons, respectively. Thinking about hard choices in terms of how different things would have to be in order for the choice to be settled not only provides an account of how hardness can vary by degree, but also opens the door to some further explanation of why it is that the choices at the harder end of the scale are so troublesome. The qualitative difference and weight of reasons that determine the degree to which one option would have to be improved to be considered better than another, when taken in combination, explain the hardness of these choices.

When an individual is ambivalent, with strong reasons for mutually exclusive alternatives, they cannot avoid choosing an option for which they have pro tanto reason against; some of their reasons will always be left unmet. The ambivalent agent cannot avoid some cause for regret, some consideration in light of which they have done the wrong thing, no matter what they choose. Moreover, the weightier one's reasons, the stronger the residual reasons that cannot be met and the more one may feel torn, or divided. "*In conflict situations,*" says Raz "*our best efforts still leave us short.*" [151] And in situations of greater conflict, our best efforts leave us even shorter.

When faced with such a choice, but the balance of reasons favours one option over the alternative, we can take consolation in the fact that we are acting as our reasons, taken together, dictate. Although we must still leave some reasons unmet, we can feel reassured that our choice is justified, safe in the knowledge that it is what any rational person would have done. However, when our reasons conflict and neither option is better supported by reasons than the other, no such consolation is forthcoming. One must make a choice that one has pro tanto reason not to make, but one cannot say why that set of reasons, rather than the other, is left unmet. This is captured by Chang's suggestion that in these choices "*our reasons run out*". [152] Part of what makes hard choices hard is that we must leave some reasons unmet without justification. The weight of reasons determines the degree to which the options would have to be different in order for the choice to be settled *and* the strength of the reasons that are left unmet.

---

[151] Raz, *From Normativity to Responsibility* p.187.
[152] Chang, "Hard Choices." p.2.

The weightier one's reasons for competing alternatives, and the greater degree of qualitative difference between them, the stronger the residual reasons that cannot satisfactorily be met and the further one is from being able to rationally justify this choice. It is this idea that leads Harry Frankfurt, in his meditation on ambivalence and wholeheartedness, to characterise ambivalence as "*a disease of the mind*", echoing Saint Augustine before him. [153] Although, strictly speaking, Frankfurt's focus is *volitional* ambivalence, rather than ambivalence in one's reasons, much of what can be said of the former applies equally well to the latter:

> "*The opportunity to act in accordance with his own inclination is a doubtful asset for an individual whose will is so divided that he is moved both to decide for a certain alterative and to decide against it. Neither of the alternatives can satisfy him, since each entails frustration of the other.*" [154]

When neither of two alternatives is better supported by reasons than the other, then being able to act in accordance with one's reasons is, likewise, a "*doubtful asset*". Either way, one will have to leave the reasons for one of these options unmet and there is no justification for failing to satisfy one set of reasons rather than the other. The hardness of a choice, thought of in terms of the degree to which things would have to be different for the choice to be settled by the balance of reasons, is directly related to the strength of the reasons that are ultimately left unmet.

# 4    Conclusion

In summary, I have argued for the following claims: (1) a meaningful distinction can be drawn between the balance and weight of reasons in practical deliberation; (2) the greater the weight of reasons underlying one's preferences, the more stable those preferences will be in response to new reasons; (3a) the weight of reasons, along with the degree of qualitative difference between alternatives, provides an account of how hard

---

[153] Frankfurt, *The Reasons of Love,* p.95
[154] Frankfurt, *Necessity, Volition, and Love*, p.102.

choices,—those between options that are on a par—vary by degree; and (3b) these factors also explain what makes those choices at the hard end of this spectrum so challenging.

# References

Anderson, Elijah. *Code of the Street: Decency, Violence, and the Moral Life of the Inner City*. WW Norton & Company, 2000.

Anscombe, Gertrude Elizabeth Margaret. "Medalist's Address: Action, Intention and 'Double Effect.'" *Proceedings of the American Catholic Philosophical Association* 56 (1982): 12–25. https://doi.org/acpaproc19825611.

Armitage, Christopher J, and Mark Conner. "Attitudinal Ambivalence: A Test of Three Key Hypotheses." *Personality and Social Psychology Bulletin* 26, no. 11 (2000): 1421–32.

Audi, Robert. *Moral Perception*. Princeton: Princeton University Press, 2013.

Ayer, Alfred Jules. *Language, Truth, and Logic*. New York: Dover Publications, 1952.

Baker, Derek. "Ambivalent Desires and the Problem with Reduction." *Philosophical Studies* 150, no. 1 (2010): 37–47.

Barnes, W H F. "A Suggestion about Value." *Analysis* 1, no. 3 (1934): 45–46.

Beddor, Bob. "Fallibility for Expressivists." *Australasian Journal of Philosophy* 98, no. 4 (2020): 763–77.

Bentham, Jeremy. *An Introduction to the Principles of Morals and Legislation (1789)*. New York: Dover Publications, 2007.

Bermúdez, José Luis. *Decision Theory and Rationality*. Oxford: Oxford University Press, 2009. https://doi.org/10.1093/acprof:oso/9780199548026.001.0001/acprof-9780199548026-chapter-3?print.

Berndsen, Mariëtte, and Joop Van der Pligt. "Ambivalence towards Meat." *Appetite* 42, no. 1 (2004): 71–78.

Bernoulli, Daniel. "Exposition of a New Theory on the Measurement of Risk (1738)." *Econometrica* 22 (1954): 23–26. https://doi.org/10.1017/S0266267115000346.

Betzler, Monika. "Sources of Practical Conflicts and Reasons for Regret." *Practical Conflicts: New Philosophical Essays*, 2004, 197–222.

Björkholm, Stina, Krister Bykvist, and Jonas Olson. "Quasi-Realism and Normative Certitude." *Synthese* 198 (2021): 794–99. https://doi.org/10.1007/s11229-020-

02553-y.

Blackburn, Simon. *Essays in Quasi-Realism*. Oxford University Press, 1993.

———. *Spreading the Word: Groundings in the Philosophy of Language*. Oxford: Clarendon Press, 1985.

———. "Truth and A Priori Possibility; Egan's Charge Against Quasi-Realism." *Australasian Journal of Philosophy* 87, no. 2 (2009): 201–13.

Bolker, Ethan. "A Simultaneous Axiomatisation of Utility and Subjective Probability." *Philosophy of Science* 34 (1967): 333–40.

———. "Functions Resembling Quotients of Measures." *Transactions of the American Mathematical Society* 124 (1966): 292–312.

Bostrom, Nick. "Pascal's Mugging." *Analysis* 69, no. 3 (2009): 443–45. https://doi.org/10.1093/analys/anp062.

Bradley, Richard. *Decision Theory With a Human Face*. Cambridge: Cambridge University Press, 2017.

Bradley, Richard, and Mareile Drechsler. "Types of Uncertainty." *Erkenntnis* 79, no. 6 (2014): 1225–48.

Broome, John. "Incommensurable Values." In *Well-Being and Morality: Essays in Honour of James Griffin*, edited by Roger Crisp and Brad Hooker. Oxford: Clarendon Press, 2000.

———. "Is Incommensurability Vagueness?" In *Incommensurability, Incomparability, and Practical Reason*, edited by Ruth Chang. Cambridge: Harvard University Press, 1997.

———. *Rationality Through Reasoning*. Oxford: Wiley-Blackwell, 2013.

———. *Weighing Goods: Equality, Uncertainty and Time*. Wiley-Blackwell, 1991.

Brown, Campbell. "The Composition of Reasons." *Synthese* 191, no. 5 (2014): 779–800. https://doi.org/10.1007/s11229-013-0299-8.

Buchak, Lara. *Risk and Rationality*. Oxford: Oxford University Press, 2013.

———. "Risk and Tradeoffs." *Erkenntnis* 79, no. S6 (2014): 1091–1117.

Burch-Brown, Joanna. "Clues for Consequentialists." *Utilitas* 26, no. 1 (2014): 105–19.

Bykvist, Krister. "Evaluative Uncertainty, Environmental Ethics, and Consequentialism."
In *Consequentialism and Environmental Ethics*, edited by Avram Hiller, Ramona Ilea,
and Leonard Kahn, 122–35. New York: Routledge, 2014.

———. "Moral Uncertainty." *Philosophy Compass* 12, no. 3 (2017).

———. "Some Critical Comments on Zimmerman's Ignorance and Moral Obligation."
*Journal of Moral Philosophy* 15 (2018): 383–400.

Bykvist, Krister, and Jonas Olson. "Against the Being For Account of Normative
Certitude." *Journal of Ethics and Social Philosophy* 6, no. 2 (2012).
https://doi.org/https://doi.org/10.26556/jesp.v6i2.63.

———. "Expressivism and Moral Certitude." *The Philosophical Quarterly* 79, no. 6 (2009):
1225–48.

———. "Non-Cognitivism and Fundamental Moral Certitude: Reply to Eriksson and
Francén Olinder." *Australasian Journal of Philosophy* 95, no. 4 (2017): 794–99.

Carlson, Erik. "Broome's Argument against Value Incomparability." *Utilitas* 16, no. 2
(2004): 220–24. https://doi.org/10.1017/S0953820804000548.

———. "Vagueness, Incomparability, and the Collapsing Principle." *Ethical Theory and
Moral Practice* 16, no. 3 (2013): 449–63. https://doi.org/10.1007/s10677-012-9352-9.

Carnap, Rudolf. *Logical Foundations of Probability*. Chicago: University of Chicago Press,
1950.

———. *The Logical Syntax of Languae*. London: Routledge, 1937.

Chang, Ruth. "Are Hard Choices Cases of Incomparability?" *Philosophical Issues* 22, no. 1
(2012): 106–26.

———. "Hard Choices." *Journal of the American Philosophical Association* 3, no. 1 (2017): 1–
21.

———. "Parity, Imprecise Comparability and the Repugnant Conclusion." *Theoria* 82,
no. 2 (2016): 182–214. https://doi.org/10.1111/theo.12096.

———. "The Possibility of Parity." *Ethics* 112, no. 4 (2002): 659–88.
https://doi.org/10.1086/339673.

———. "The Possibility of Parity." *Ethics* 112, no. 4 (2002): 659–88.

Conner, Mark, and Christopher J Armitage. "Attitudinal Ambivalence." In *Attitudes and Attitude Change*, edited by W Crano and R Prislin, 261–86. New York: Psychology Press, 2008.

Conner, Mark, Paul Sparks, Rachel Povey, Rhiannon James, Richard Shepherd, and Christopher J Armitage. "Moderator Effects of Attitudinal Ambivalence on Attitude–Behaviour Relationships." *European Journal of Social Psychology* 32, no. 5 (2002): 705–18.

Constantinescu, Cristian. "Value Incomparability and Indeterminacy." *Ethical Theory and Moral Practice* 15, no. 1 (2012): 57–70.

Cowen, Tyler. "The Epistemic Problem Does Not Refute Consequentialism." *Utilitas* 18, no. 4 (2006): 383–99. https://doi.org/10.1017/S0953820806002172.

Crisp, Roger. "Review of Kupperman, Value... and What Follows." *Philosophy* 75 (2000): 458–92.

Cullison, Andrew. "Moral Perception." *European Journal of Philosophy* 18, no. 2 (2010): 159–75.

Cullity, Garrett. "Weighing Reasons." In *The Oxford Handbook of Reasons and Normativity*, edited by Daniel Star, 423–42. Oxford: Oxford University Press, 2018. https://doi.org/10.1093/oxfordhb/9780199657889.013.19.

Dancy, Jonathan. *Ethics Without Principles*. Oxford: Oxford University Press, 2004.

Davidson, Donald. "How Is Weakness of The Will Possible?" In *Essays on Actions and Events*, 21–42. Oxford: Clarendon Press, n.d.

Dietrich, Franz, and Christian List. "Mentalism Versus Behaviourism in Economics: A Philosophy-of-Science Perspective." *Economics and Philosophy* 32, no. 2 (2016): 249–81.

Disney, Walt, Ben Sharpsteen, and Hamilton Luske. *Pinocchio*. RKO Radio Pictures, 1940.

Dorsey, Dale. "Consequentialism, Metaphysical Realism and the Argument from Cluelessness." *The Philosophical Quarterly* 62, no. 246 (2012): 48–70. https://doi.org/10.1111/j.1467-9213.2011.713.x.

Egan, Andy. "Quasi-Realism and Fundamental Moral Error." *Australasian Journal of Philosophy* 85, no. 2 (2007): 205–19.

Eriksson, John, and Ragnar Francén Olinder. "Non-Cognitivism and the Classification Account of Moral Uncertainty." *Australasian Journal of Philosophy* 94, no. 4 (2016): 719–35.

Feldman, Fred. "Actual Utility, The Objection from Impracticality, and the Move to Expected Utility." *Philosophical Studies* 129, no. 1 (2006): 49–79. https://doi.org/10.1007/s11098-005-3021-y.

Feldman, S D, and A Hazlett. "In Defense of Ambivalence." manuscript, n.d.

Finlay, Stephen. *Confusion of Tongues: A Theory of Normative Language*. Oxford: Oxford University Press, 2014.

Fiske, Alan Page, and Tage Shakti Rai. *Virtuous Violence: Hurting and Killing to Create, Sustain, End, and Honor Social Relationships*. Cambridge: Cambridge University Press, 2015.

Foot, Philippa. "The Problem of Abortion and the Doctrine of Double Effect." *Oxford Review* 5 (1967): 5–15.

Frankfurt, Harry. *Necessity, Volition, and Love*. Cambridge University Press, 1999.

———. *The Reasons of Love*. Princeton: Princeton University Press, 2004.

Frankfurt, Harry G. *The Importance of What We Care About*. Cambridge University Press, 1988.

Gärdenfors, Peter, and Nils-Eric Sahlin. "Unreliable Probabilities, Risk Taking, and Decision Making." *Synthese* 53, no. 3 (1982): 361–86.

Gibbard, Allan. *Thinking How to Live*. Harvard University Press, 2003.

———. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Harvard University Press, 1990.

Gill, Michael B. "Metaethical Variability, Incoherence, and Error." In *Moral Psychology*, edited by Walter Sinnott-Armstrong, 387–402. Cambridge: MIT Press, 2008.

GiveWell. "Our Top Charities," November 2020. https://www.givewell.org/charities/top-charities.

Gracely, Edward J. "On the Noncomparability of Judgments Made by Different Ethical Theories." *Metaphilosophy* 27, no. 3 (1996): 327–32.

Graham, Peter A. "In Defense of Objectivism about Moral Obligation." *Ethics* 121, no. 1 (2010): 88–115. https://doi.org/10.1086/656328.

Greaves, Hilary. "Cluelessness." *Proceedings of the Aristotelian Society* 116, no. 3 (2016): 311–39. https://doi.org/10.1093/arisoc/aow018.

Greenspan, Patricia. "A Case of Mixed Feelings: Ambivalence and the Logic of Emotion." In *Explaining Emotions*, edited by Amelie Oksenberg Rorty, 223–50. Berkeley: University of California Press, 1980.

Gustafsson, Johan E, and Olle Torpman. "In Defence of My Favourite Theory." *Pacific Philosophical Quarterly* 95, no. 2 (2014): 159–74.

Hare, Caspar. "Take the Sugar." *Analysis* 70, no. 2 (2010): 237–47. https://doi.org/10.1093/analys/anp174.

Hare, Richard Mervyn. *The Language of Morals*. Oxford: Clarendon Press, 1952.

Harman, Elizabeth. "The Irrelevance of Moral Uncertainty." In *Oxford Studies in Metaethics, Volume 10*, edited by Russ Shafer-Landau, 10:53–79. Oxford University Press Oxford, 2015.

Hausman, Daniel. *Preference, Value, Choice, and Welfare*. Cambridge: Cambridge University Press, 2011.

———. "Sympathy, Commitment, and Preference." *Economics and Philosophy* 21, no. 1 (2005): 33–50.

Hedden, Brian. "Does MITE Make Right? On Decision-Making under Normative Uncertainty." In *Oxford Studies in Metaethics Volume 11*, edited by Russ Schafer-Landau, 102–28. Oxford: Oxford University Press, 2016.

Hieronymi, P. "The Wrong Kind of Reason." *Journal of Philosophy* 102, no. 9 (2005): 437–57. https://doi.org/10.5840/jphil2005102933.

Hudson, James Leonard. "Subjectivization in Ethics." *American Philosophical Quarterly* 26, no. 3 (1989): 221–29.

Hume, David. *A Treatise of Human Nature (1739)*. Edited by David F. Norton and Mary J. Norton. Oxford: Clarendon Press, 2007.

Hyman, John. *Action, Knowledge, and Will*. Oxford: Oxford University Press, 2015.

Jackson, Frank. "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection." *Ethics* 101, no. 3 (1991): 461–82.

———. "Internal Conflicts in Desires and Morals." *American Philosophical Quarterly* 22, no. 2 (1985): 105–14. https://doi.org/10.2307/20014086.

———. "Non-Cognitivism, Validity and Conditionals." In *Dale Jamieson (Ed.) Singer and His Critics*, 18–37. Oxford: Blackwell, 1999.

Jacobson, Daniel. "Fitting Attitude Theories of Value (Stanford Encyclopedia of Philosophy/Spring 2011 Edition)," 2011. https://plato.stanford.edu/archives/spr2011/entries/fitting-attitude-theories/.

Jeffrey, Richard. *The Logic of Decision*. 3rd ed. Chicago: University of Chicago Press, 1965.

Jerzak, Ethan. "Two Ways to Want?" *The Journal of Philosophy* 116, no. 2 (2019): 65–98.

Johnson King, Zoë. "Accidentally Doing the Right Thing." *Philosophy and Phenomenological Research* 100, no. 1 (2020): 186–206.

Joyce, James. *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press, 1999.

Kant, Immanuel. *Groundwork of the Metaphysics of Morals (1785)*. Edited by Mary Gregor and Jens Timmerman. Cambridge: Cambridge University Press, 2012.

Kaplan, Kalman J. "On the Ambivalence-Indifference Problem in Attitude Theory and Measurement: A Suggested Modification of the Semantic Differential Technique." *Psychological Bulletin* 77, no. 5 (1972): 361.

Kearns, Stephen, and Daniel Star. "Reasons: Explanations or Evidence?" *Ethics* 119, no. 1 (October 2008): 31–56. https://doi.org/10.1086/592587.

———. "Reasons as Evidence." In *Reasons as Evidence*, edited by Russ Shafer-Landau, 4:215–42. Oxford: Oxford University Press, 2009.

Keynes, John Maynard. *A Treatise on Probability*. London: Macmillan, 1921.

Klein, Peter. "The Virtues of Inconsistency." *The Monist* 68, no. 1 (1985): 105–35.

Köhler, Sebastian. "What Is the Problem with Fundamental Moral Error?" *Australasian Journal of Philosophy* 93, no. 1 (2015): 161–65.

Korsgaard, Christine. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press, 2009.

———. *The Sources of Normativity*. Cambridge: Cambridge University Press, 1996.

Kristjánsson, Kristján. "The Trouble with Ambivalent Emotions." *Philosophy* 85, no. 4 (2010): 485–510.

Lam, Garrett. "Making Quasi-Realists Admit of Fundamental Moral Fallibility." *Australasian Journal of Philosophy* 98, no. 2 (2020): 294–303.

Lee, Spike. *Do The Right Thing*. Universal Pictures, 1989.

Lenman, James. "Consequentialism and Cluelessness." *Philosophy & Public Affairs* 29, no. 4 (2000): 342–70. https://doi.org/10.1111/j.1088-4963.2000.00342.x.

———. "Non-Cognitivism and the Dimensions of Evaluative Judgement." *Brown Electronic Article Review Service (BEARS)*, 2003.

Levi, Isaac. *Hard Choices*. Cambridge: Cambridge University Press, 1986.

———. *The Enterprise of Knowledge*. Cambridge: MIT Press, 1980.

———. "The Weight of Argument." In *Fundamental Uncertainty: Rationality and Plausible Reasoning*, edited by S. M. D. Brandolini and R. Scazziere, 39–58. Basingstoke: Palgrave Macmillan, 2011.

Lewis, David. "Radical Interpretation." *Synthese* 27, no. 3 (1974): 331–44.

Lipkus, Isaac M, Jeffrey D Green, John R Feaganes, and Constantine Sedikides. "The Relationship Between Attitudinal Ambivalence and Desire to Quit Smoking Among College Smokers." *Journal of Applied Social Psychology* 31, no. 1 (2001): 113–33.

Lockhart, Ted. *Moral Uncertainty and Its Consequences*. Oxford: Oxford University Press, 2000.

Lord, Errol, and Barry Maguire. "An Opinionated Guide to the Weight of Reasons." In *Weighing Reasons*, 3–24. Oxford: Oxford University Press, 2016. https://doi.org/10.1093/acprof:oso/9780199315192.003.0001.

MacAskill, William. "Normative Uncertainty." DPhil thesis, University of Oxford, 2014.

MacAskill, William, Krister Bykvist, and Toby Ord. *Moral Uncertainty*. Oxford: Oxford

University Press, 2020.

MacAskill, William, and Toby Ord. "Why Maximize Expected Choice-Worthiness?" *Noûs* 54, no. 2 (2020): 327–53.

Mackie, John Leslie. *Ethics: Inventing Right and Wrong*. New York: Penguin Books, 1977.

Maguire, Barry, and Justin Snedegar. "Normative Metaphysics for Accountants." *Philosophical Studies* 178, no. 2 (February 1, 2021): 363–84. https://doi.org/10.1007/s11098-020-01435-w.

Makinson, David C. "The Paradox of the Preface." *Analysis* 25, no. 6 (1965): 205–7.

McHugh, Conor, and Jonathan Way. "Fittingness First." *Ethics* 126, no. 3 (April 1, 2016): 575–606. https://doi.org/10.1086/684712.

Mill, John Stewart. *Utilitarianism (1861)*. Edited by Roger Crisp. Oxford: Oxford Univserity Press, 1998.

Moran, Richard. "Review Essay on 'The Reasons of Love.'" *Philosophy and Phenomenological Research* 74, no. 2 (2007): 463–75.

Neumann, John von, and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton: Princeton University Press, 1944.

Nissan-Rozen, Ittay. "Against Moral Hedging." *Economics and Philosophy*, no. 3 (2015).

Norcross, Alastair. "Consequentialism and the Unforeseeable Future." *Analysis* 50, no. 4 (1990): 253–56. https://doi.org/10.1093/analys/50.4.253.

Nussbaum, Martha. "Aeschylus and Practical Conflict." *Ethics* 95, no. 2 (1985): 233–67.

Okasha, Samir. "On the Interpretation of Decision Theory." *Economics and Philosophy* 32, no. 3 (2016): 409–33.

Parfit, Derek. "Can We Avoid the Repugnant Conclusion?" *Theoria* 82, no. 2 (2016): 110–27. https://doi.org/10.1111/theo.12097.

———. "Rationality and Reasons." In *Exploring Practical Philosophy*, edited by Dan Egonsson, Björn Petersson, Jonas Josefsson, and Toni Rønnow-Rasmussen, 17–41. Aldershot: Ashgate, 2001.

———. *Reasons and Persons*. Oxford University Press, 1984.

Parfit, Derek, and John Broome. "Reasons and Motivation." *Proceedings of the Aristotelian Society, Supplementary Volumes* 71 (1997): 99–146.

Peirce, Charles Sanders. "The Probability of Induction." In *The Essential Peirce: Selected Philosophical Writings. Volume 1 (1867–1893)*, edited by Nathan Houser and Christian Kloesel, 155–69. Bloomington: Indiana University Press, 1992.

Pettit, Philip. "Construing Sen on Commitment." *Economics and Philosophy* 21, no. 1 (2005): 15–32. https://doi.org/10.1017/S0266267104000367.

———. "Decision Theory and Folk Psychology." In *Essays in the Foundations of Decision Theory*, edited by Michael Bacharach and Susan Hurley, 147–75. Oxford: Blackwell, 1991.

Popper, Karl. *The Logic of Scientific Discovery (1959)*. London: Routledge Classics, 2002.

Priester, Joseph R, and Richard E Petty. "Extending the Bases of Subjective Attitudinal Ambivalence: Interpersonal and Intrapersonal Antecedents of Evaluative Tension." *Journal of Personality and Social Psychology* 80, no. 1 (2001): 19.

Prinz, Jesse J. *Gut Reactions: A Perceptual Theory of Emotion*. Oxford: Oxford University Press, 2004.

Railton, Peter. "Consequentialism , and the Demands of Morality." *Philosophy & Public Affairs* 13, no. 2 (1984): 134–71. https://doi.org/10.2307/2265273.

Ramsey, Frank Plumpton. "Truth and Probability." In *Philosophical Papers*, 52–109. Cambridge: Cambridge University Press, 1990.

Raz, Joseph. *Engaging Reasons*. Oxford: Oxford University Press, 1999.

———. *From Normativity to Responsibility*. Oxford University Press, 2011.

———. *Practical Reason and Norms*. Oxford: Oxford University Press, 1975.

———. *The Morality of Freedom*. Clarendon Press, 1986.

———. *The Morality of Freedom*. Oxford: Clarendon Press, 1986.

Razinsky, Hili. "An Outline for Ambivalence of Value Judgment." *The Journal of Value Inquiry* 48, no. 3 (2014): 469–88.

Ridge, Michael. "Certitude, Importance, and Robustness for Non-Cognitivists." *Brown*

*Electronic Article Review Service (BEARS)*, 2003.

———. "Ecumenical Expressivism: Finessing Frege." *Ethics* 116, no. 2 (2006): 302–36.

Rosenthal, Chelsea. "What Decision Theory Can't Tell Us About Moral Uncertainty." *Philosophical Studies*, 2020. https://doi.org/https://doi.org/10.1007/s11098-020-01571-3.

Ross, Jacob. "Rejecting Ethical Deflationism." *Ethics* 116, no. 4 (2006): 742–68.

Ross, William David. *The Right and The Good*. Oxford: Oxford University Press, 1930.

Sartre, Jean-Paul. *Existentialism Is a Humanism*. Yale University Press, 2007.

Savage, Leonard. *The Foundations of Statistics*. New York: Dover, 1954.

Scanlon, Thomas. *What We Owe to Each Other*. Cambridge: Harvard University Press, 1998.

Schroeder, Mark. *Being For*. Oxford: Oxford University Press, 2008.

———. "Expression for Expressivists." *Philosophy and Phenomenological Research* 76, no. 1 (2008): 86–116.

Sen, Amartya. "Rational Fools: A Critique of the Behavioural Foundations of Economic Theory." *Philosoph & Public Affairs* 6, no. 4 (1977): 317–44.

———. *Rationality and Freedom*. Cambridge: Hardvard University Press, 2002.

Sepielli, Andrew. "'Along an Imperfectly-Lighted Path': Practical Rationality and Normative Uncertainty." Doctoral Thesis, Rutgers University, 2010.

———. "Decision-Making Under Moral Uncertainty." In *Routledge Handbook on Moral Epistemology*, edited by Aaron Zimmerman, Karen Jones, and Mark Timmons, 508–21. New York: Routledge, 2019.

———. "Normative Uncertainty for Non-Cognitivists." *Philosophical Studies* 160, no. 2 (2012): 191–207.

———. "Subjective and Objective Reasons." In *The Oxford Handbook of Reasons and Normativity*, edited by Daniel Star, 784–99. Oxford: Oxford University Press, 2018.

———. "What To Do When You Don't Know What To Do." In *Oxford Studies in Metaethics, Volume IV*, edited by Russ Shafer-Landau, 4:5–28. Oxford: Oxford

University Press, 2009.

———. "What to Do When You Don't Know What to Do When You Don't Know What to Do …" *Noûs* 48, no. 3 (2014): 521–44.

Setiya, Kieran. "What Is a Reason to Act?" *Philosophical Studies* 167, no. 2 (2014): 221–35. https://doi.org/10.1007/s11098-012-0086-2.

Silverstein, Matthew. "Reducing Reasons." *Journal of Ethics and Social Philosophy* 10, no. 1 (June 7, 2017): 1–23. https://doi.org/10.26556/jesp.v10i1.95.

Singer, Peter. "Famine, Affluence, and Morality." *Philosophy and Public Affairs* 1, no. 3 (1972): 229–43.

Sinnott-Armstrong, Walter. "Mixed Up Meta-Ethics." *Philosophical Issues* 19, no. 1 (2009): 235–56.

Skyrms, Brian. "Resiliency, Propensities, and Causal Necessity." *The Journal of Philosophy* 74, no. 11 (1977): 704–13.

Slote, Michael A. *Beyond Optimizing: A Study of Rational Choice*. Cambridge: Harvard University Press, 1989.

Slote, Michael, and Philip Pettit. "Satisficing Consequentialism." *Proceedings of the Aristotelian Society* 58 (1984): 139–63.

Smith, Michael. "A Constitutivist Theory of Reasons: Its Promise and Parts." *Law, Ethics and Philosophy* 1 (2013): 1–30.

———. "Evaluation, Uncertainty and Motivation." *Ethical Theory and Moral Practice* 5, no. 3 (2002): 305–20.

———. "The Magic of Constitutivism." *American Philosophical Quarterly* 52, no. 2 (2015): 187–200.

———. *The Moral Problem*. Oxford: Blackwell, 1994.

Snedegar, Justin. "Reasons For and Reasons Against." *Philosophical Studies* 175, no. 3 (2018): 725–43. https://doi.org/10.1007/s11098-017-0889-2.

Spinoza, Benedict de. "Ethics." In *The Collected Works of Spinoza Volume I*, edited by Edwin Curley, Vol. 1. Princeton: Princeton University Press, 1996.

Staffel, Julia. "Expressivism, Normative Uncertainty, and Arguments for Probabilism."
In *Oxford Studies in Epistemology, Volume 6*, edited by Tamar Szabó Gendler and John
Hawthorne, 161–89. Oxford: Oxford University Press, 2019.

Stevenson, Charles Leslie. "The Emotive Meaning of Ethical Terms." *Mind* 46, no. 181
(1937): 14–31.

Strawson, Peter. "Freedom and Resentment." *Proceedings of the British Academy* 48 (1962):
25. https://doi.org/10.1073/PNAS.48.1.1.

Swindell, Jennifer S. "Ambivalence." *Philosophical Explorations* 13, no. 1 (2010): 23–34.

Tarsney, Christian. "Intertheoretic Value Comparison: A Modest Proposal." *Journal of
Moral Philosophy* 15, no. 3 (2018): 324–44.

Temkin, Larry S. *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning.*
Oxford: Oxford University Press, 2012.

Thoma, Johanna. "Advice for the Steady: Decision Theory and the Requirements of
Instrumental Rationality." PhD Thesis, University of Toronto, 2017.

———. "Decision Theory." In *The Open Handbook of Formal Epistemology*, edited by
Richard Pettigrew and Jonathan Weisberg, 57–106. PhilPapers Foundation, 2019.

———. "Judgementalism About Normative Decision Theory." *Synthese*, n.d.
https://doi.org/10.1007/s11229-019-02487-0.

Thomas, William Isaac, and Florian Znaniecki. *The Polish Peasant in Europe and America:
Monograph of an Immigrant Group.* Chicago: University of Illinois Press, 1918.

Thompson, Megan M, Mark P Zanna, and Dale W Griffin. "Let's Not Be Indifferent
about (Attitudinal) Ambivalence." In *Attitude Strength: Antecedents and Consequences*,
edited by R. E. Petty and J. A. Krosnick, 4:361–86. New York: Lawrence Erlbaum
Associates, Inc., 1995.

Thomson, Judith Jarvis. *Normativity.* Chicago: Open Court, 2008.

Toulmin, Stephen. *En Examination of the Place of Reason in Ethics.* Cambridge: Cambridge
University Press, 1950.

Tucker, Chris. "The Dual Scale Model of Weighing Reasons." *Nous*, 2021.
https://doi.org/10.1111/nous.12361.

Velleman, David. *How We Get Along*. Cambridge: Cambridge University Press, 2009.

———. *Self to Self*. New York: Cambridge University Press, 2005.

Way, Jonathan. "Reasons as Premises of Good Reasoning." *Pacific Philosophical Quarterly* 98, no. 2 (2017): 251–70. https://doi.org/10.1111/papq.12135.

Weatherson, Brian. "Keynes, Uncertainty and Interest Rates." *Cambridge Journal of Economics* 26, no. 1 (2002): 47–62.

———. *Normative Externalism*. Oxford: Oxford University Press, 2019.

———. "Running Risks Morally." *Philosophical Studies* 167, no. 1 (2014): 141–63.

Whiting, Daniel. "Right in Some Respects: Reasons as Evidence." *Philosophical Studies* 175, no. 9 (2018): 2191–2208. https://doi.org/10.1007/s11098-017-0954-x.

Williams, Bernard. *Ethics and the Limits of Philosophy*. Abingdon: Routledge, 1985.

Wright, Jennifer Cole. "The Fact and Function of Meta-Ethical Pluralism: Exploring the Evidence." In *Oxford Studies in Experimental Philosophy, Volume 2*, edited by Tania Lombrozo, Joshua Knobe, and Shaun Nichols, 119–50. Oxford: Oxford University Press, 2018.

Wringe, Bill. "Ambivalence for Cognitivists: A Lesson from Chrysippus?" *Thought: A Journal of Philosophy* 6, no. 3 (2017): 147–56.

Yudkowsky, Eliezer. "Pascal's Mugging: Tiny Probabilities of Vast Utilities." LessWrong.com, 2007. https://www.lesswrong.com/posts/a5JAiTdytou3Jg749/pascal-s-mugging-tiny-probabilities-of-vast-utilities.

Zimmerman, Michael J. "Is Moral Obligation Objective or Subjective?" *Utilitas* 18, no. 4 (2006): 329–61. https://doi.org/10.1017/S0953820806002159.